

(43) International Publication Date 28 June 2001 (28.06.2001)

PCT

(10) International Publication Number WO 01/45748 A1

(51) International Patent Classification7: A61K 48/00, C12Q 1/70

NJ 07065-0907 (US). FU, Tong-Ming [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US).

(21) International Application Number: PCT/US00/34724

(74) Common Representative: MERCK & CO., INC.: 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US).

(22) International Filing Date:

21 December 2000 (21.12.2000)

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, ITU, ID, IL, IN, IS, JP, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data: 60/171,542 22 December 1999 (22.12.1999)

(71) Applicant (for all designated States except US): MERCK & CO., INC. [US/US]; 126 East Lincoln Avenue, Rahway,

NJ 07065-0907 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): SHIVER, John, W. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). PERRY, Helen, C. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). CASIMIRO, Danilo, R. [PH/US]; 126 East Lincoln Avenue, Rahway, (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

With international search report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: POLYNUCLEOTIDE VACCINES EXPRESSING CODON OPTIMIZED HIV-1 POL AND MODIFIED HIV-1 POL

(57) Abstract: Pharmaceutical compositions which comprise HIV Pol DNA vaccines are disclosed, along with the production and use of these DNA vaccines. The pol-based DNA vaccines of the invention are administered directly introduced into living vertebrate activity, RNase H activity and integrase activity, inducing a cellular immune response which specifically recognizes human immunodeficiency virus-1 (HIV-1). The DNA molecules which comprise the open reading frame of these DNA vaccines are synthetic DNA molecules encoding codon optimized HIV-1 Pol and codon optimized inactive derivatives of optimized HIV-1 Pol, including DNA molecules which encode inactive Pol proteins which comprise an amino terminal leader peptide.

TITLE OF THE INVENTION
POLYNUCLEOTIDE VACCINES EXPRESSING CODON OPTIMIZED HIV-1
POL AND MODIFIED HIV-1 POL

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit, under 35 U.S.C. §119(e), of U.S. provisional application 60/171,542, filed December 22, 1999.

10

20

25

30

5

STATEMENT REGARDING FEDERALLY-SPONSORED R&D Not Applicable

15 REFERENCE TO MICROFICHE APPENDIX Not Applicable

FIELD OF THE INVENTION

The present invention relates to HIV Pol polynucleotide pharmaceutical products, as well as the production and use thereof which, when directly introduced into living vertebrate tissue, preferably a mammalian host such as a human or a non-human mammal of commercial or domestic veterinary importance, express the HIV Pol protein or biologically relevant portions thereof within the animal, inducing a cellular immune response which specifically recognizes human immunodeficiency virus-1 (HIV-1). The polynucleotides of the present invention are synthetic DNA molecules encoding codon optimized HIV-1 Pol and derivatives of optimized HIV-1 Pol, including constructs wherein protease, reverse transcriptase, RNAse H and integrase activity of HIV-1 Pol is inactivated. The polynucleotide vaccines of the present invention should offer a prophylactic advantage to previously uninfected individuals and/or provide a therapeutic effect by reducing viral load levels within an infected individual, thus prolonging the asymptomatic phase of HIV-1 infection.

BACKGROUND OF THE INVENTION

5

10

15

20

25

30

Human Immunodeficiency Virus-1 (HIV-1) is the etiological agent of acquired human immune deficiency syndrome (AIDS) and related disorders. HIV-1 is an RNA virus of the Retroviridae family and exhibits the 5'LTR-gag-pol-env-LTR 3'organization of all retroviruses. The integrated form of HIV-1, known as the provirus, is approximately 9.8 Kb in length. Each end of the viral genome contains flanking sequences known as long terminal repeats (LTRs). The HIV genes encode at least nine proteins and are divided into three classes; the major structural proteins (Gag, Pol, and Env), the regulatory proteins (Tat and Rev); and the accessory proteins (Vpu, Vpr, Vif and Nef).

The gag gene encodes a 55-kilodalton (kDa) precursor protein (p55) which is expressed from the unspliced viral mRNA and is proteolytically processed by the HIV protease, a product of the *pol* gene. The mature p55 protein products are p17 (matrix), p24 (capsid), p9 (nucleocapsid) and p6.

The *pol* gene encodes proteins necessary for virus replication; a reverse transcriptase, a protease, integrase and RNAse H. These viral proteins are expressed as a Gag-Pol fusion protein, a 160 kDa precursor protein which is generated via a ribosomal frame shifting. The viral encoded protease proteolytically cleaves the Pol polypeptide away from the Gag-Pol fusion and further cleaves the Pol polypeptide to the mature proteins which provide protease (Pro, P10), reverse transcriptase (RT, P50), integrase (IN, p31) and RNAse H (RNAse, p15) activities.

The *nef* gene encodes an early accessory HIV protein (Nef) which has been shown to possess several activities such as down regulating CD4 expression, disturbing T-cell activation and stimulating HIV infectivity.

The env gene encodes the viral envelope glycoprotein that is translated as a 160-kilodalton (kDa) precursor (gp160) and then cleaved by a cellular protease to yield the external 120-kDa envelope glycoprotein (gp120) and the transmembrane 41-kDa envelope glycoprotein (gp41). Gp120 and gp41 remain associated and are displayed on the viral particles and the surface of HIV-infected cells.

The *tat* gene encodes a long form and a short form of the Tat protein, a RNA binding protein which is a transcriptional transactivator essential for HIV-1 replication.

The *rev* gene encodes the 13 kDa Rev protein, a RNA binding protein. The Rev protein binds to a region of the viral RNA termed the Rev response element

PCT/US00/34724 WO 01/45748

(RRE). The Rev protein is promotes transfer of unspliced viral RNA from the nucleus to the cytoplasm. The Rev protein is required for HIV late gene expression and in turn, HIV replication.

5

10

15

25

30

Gp120 binds to the CD4/chemokine receptor present on the surface of helper T-lymphocytes, macrophages and other target cells in addition to other co-receptor molecules. X4 (macrophage tropic) virus show tropism for CD4/CXCR4 complexes while a R5 (T-cell line tropic) virus interacts with a CD4/CCR5 receptor complex. After gp120 binds to CD4, gp41 mediates the fusion event responsible for virus entry. The virus fuses with and enters the target cell, followed by reverse transcription of its single stranded RNA genome into the double-stranded DNA via a RNA dependent DNA polymerase. The viral DNA, known as provirus, enters the cell nucleus, where the viral DNA directs the production of new viral RNA within the nucleus, expression of early and late HIV viral proteins, and subsequently the production and cellular release of new virus particles. Recent advances in the ability to detect viral load within the host shows that the primary infection results in an extremely high generation and tissue distribution of the virus, followed by a steady state level of virus (albeit through a continual viral production and turnover during this phase), leading ultimately to another burst of virus load which leads to the onset of clinical AIDS. Productively infected cells have a half life of several days, whereas chronically or latently infected cells have a 3-week half life, followed by non-productively infected 20 cells which have a long half life (over 100 days) but do not significantly contribute to day to day viral loads seen throughout the course of disease.

Destruction of CD4 helper T lymphocytes, which are critical to immune defense, is a major cause of the progressive immune dysfunction that is the hallmark of HIV infection. The loss of CD4 T-cells seriously impairs the body's ability to fight most invaders, but it has a particularly severe impact on the defenses against viruses, fungi, parasites and certain bacteria, including mycobacteria.

Effective treatment regimens for HIV-1 infected individuals have become available recently. However, these drugs will not have a significant impact on the disease in many parts of the world and they will have a minimal impact in halting the spread of infection within the human population. As is true of many other infectious diseases, a significant epidemiologic impact on the spread of HIV-1 infection will only occur subsequent to the development and introduction of an effective vaccine. There are a number of factors that have contributed to the lack of successful vaccine

5

10

15

20

25

30

development to date. As noted above, it is now apparent that in a chronically infected person there exists constant virus production in spite of the presence of anti-HIV-1 humoral and cellular immune responses and destruction of virally infected cells. As in the case of other infectious diseases, the outcome of disease is the result of a balance between the kinetics and the magnitude of the immune response and the pathogen replicative rate and accessibility to the immune response. Pre-existing immunity may be more successful with an acute infection than an evolving immune response can be with an established infection. A second factor is the considerable genetic variability of the virus. Although anti-HIV-1 antibodies exist that can neutralize HIV-1 infectivity in cell culture, these antibodies are generally virus isolate-specific in their activity. It has proven impossible to define serological groupings of HIV-1 using traditional methods. Rather, the virus seems to define a serological "continuum" so that individual neutralizing antibody responses, at best, are effective against only a handful of viral variants. Given this latter observation, it would be useful to identify immunogens and related delivery technologies that are likely to elicit anti-HIV-1 cellular immune responses. It is known that in order to generate CTL responses antigen must be synthesized within or introduced into cells. subsequently processed into small peptides by the proteasome complex, and translocated into the endoplasmic reticulum/Golgi complex secretory pathway for eventual association with major histocompatibility complex (MHC) class I proteins. CD8⁺ T lymphocytes recognize antigen in association with class I MHC via the T cell receptor (TCR) and the CD8 cell surface protein. Activation of naive CD8⁺ T cells into activated effector or memory cells generally requires both TCR engagement of antigen as described above as well as engagement of costimulatory proteins. Optimal induction of CTL responses usually requires "help" in the form of cytokines from CD4+T lymphocytes which recognize antigen associated with MHC class II molecules via TCR and CD4 engagement.

Larder, et al., (1987, *Nature* 327: 716-717) and Larder, et al., (1989, *Proc. Natl. Acad. Sci.* 86: 4803-4807) disclose site specific mutagenesis of HIV-1 RT and the effect such changes have on *in vitro* activity and infectivity related to interaction with known inhibitors of RT.

Davies, et al. (1991, *Science* 252:, 88-95) disclose the crystal structure of the RNase H domain of HIV-1 Pol.

Schatz, et al. (1989, FEBS Lett. 257: 311-314) disclose that mutations Glu478Gln and His539Phe in a complete HIV-1 RT/RNase H DNA fragment results in defective RNase activity without effecting RT activity.

Mizrahi, et al. (1990, *Nucl. Acids. Res.* 18: pp. 5359-5353) disclose additional mutations Asp443Asn and Asp498Asn in the RNase region of the *pol* gene which also results in defective RNase activity. The authors note that the Asp498Asn mutant was difficult to characterize due to instability of this mutant protein.

Leavitt, et al. (1993, J. Biol. Chem. 268: 2113-2119) disclose several mutations, including a Asp64Val mutation, which show differing effect on HIV-1 integrase (IN) activity.

Wiskerchen, et al. (1995, *J. Virol*. 69: 376-386) disclose singe and double mutants, including mutation of aspartic acid residues which effect HIV-1 IN and viral replication functions.

It would be of great import in the battle against AIDS to produce a prophylactic- and/or therapeutic-based HIV vaccine which generates a strong cellular immune response against an HIV infection. The present invention addresses and meets this needs by disclosing a class of DNA vaccines based on host delivery and expression of modified versions of the HIV-1 gene, *pol*.

20 SUMMARY OF THE INVENTION

5

10

25

30

The present invention relates to synthetic DNA molecules (also referred to herein as "polynucleotides") and associated DNA vaccines (also referred to herein as "polynucleotide vaccines") which elicit cellular immune and humoral responses upon administration to the host, including primates and especially humans, and also including a non-human mammal of commercial or domestic veterinary importance. An effect of the cellular immune-directed vaccines of the present invention should be the lower transmission rate to previously uninfected individuals and/or reduction in the levels of the viral loads within an infected individual, so as to prolong the asymptomatic phase of HIV-1 infection. In particular, the present invention relates to DNA vaccines which encode various forms of HIV-1 Pol, wherein administration, intracellular delivery and expression of the HIV-1 Pol gene of interest elicits a host CTL and Th response. The preferred synthetic DNA molecules of the present invention encode codon optimized versions of wild type HIV-1 Pol, codon optimized versions of HIV-1 Pol fusion proteins, and codon optimized versions of HIV-1 Pol

proteins and fusion protein, including but not limited to *pol* modifications involving residues within the catalytic regions responsible for RT, RNase and IN activity within the host cell.

A particular embodiment of the present invention relates to codon optimized wt-pol DNA constructs wherein DNA sequences encoding the protease (PR) activity are deleted, leaving codon optimized "wild type" sequences which encode RT (reverse transcriptase and RNase H activity) and IN integrase activity. The nucleotide sequence of a DNA molecule which encodes this protein is disclosed herein as SEQ ID NO:1 and the corresponding amino acid sequence of the expressed protein is disclosed herein as SEQ ID NO:2.

5

10

15

20

25

30

The present invention preferably relates to a HIV-1 DNA pol construct which is devoid of DNA sequences encoding any PR activity, as well as containing a mutation(s) which at least partially, and preferably substantially, abolishes RT, RNase and/or IN activity. One type of HIV-1 pol mutant may include but is not limited to a mutated DNA molecule comprising at least one nucleotide substitution which results in a point mutation which effectively alters an active site within the RT, RNase and/or IN regions of the expressed protein, resulting in at least substantially decreased enzymatic activity for the RT, RNase H and/or IN functions of HIV-1 Pol. In a preferred embodiment of this portion of the invention, a HIV-1 DNA pol construct contains a mutation or mutations within the Pol coding region which effectively abolishes RT, RNase H and IN activity. An especially preferable HIV-1 DNA pol construct in a DNA molecule which contains at least one point mutation which alters the active site of the RT, RNase H and IN domains of Pol, such that each activity is at least substantially abolished. Such a HTV-1 Pol mutant will most likely comprise at least one point mutation in or around each catalytic domain responsible for RT, RNase H and IN activity, respectfully. To this end, an especially preferred HIV-1 DNA pol construct is exemplified herein and contains nine codon substitution mutations which results in an inactivated Pol protein (IA Pol: SEQ ID NO:4, Figure 2A-C) which has no PR, RT, RNase or IN activity, wherein three such point mutations reside within each of the RT, RNase and IN catalytic domains. Any combination of the mutations disclosed herein may suitable and therefore may be utilized as an IA-Pol-based vaccine of the present invention. While addition and deletion mutations are contemplated and within the scope of the invention, the

PCT/US00/34724 WO 01/45748

preferred mutation is a point mutation resulting in a substitution of the wild type amino acid with an alternative amino acid residue.

5

10

15

Another aspect of the present invention is to generate HIV-1 Pol-based vaccine constructions which comprise a eukaryotic trafficking signal peptide such as the leader peptide from human tPA. To this end, the present invention relates to a DNA molecule which encodes a codon optimized wt-pol DNA construct wherein the protease (PR) activity is deleted and a human tPA leader sequence is fused to the 5' end of the coding region. A DNA molecule which encodes this protein is disclosed herein as SEQ ID NO:5, the open reading frame disclosed herein as SEQ ID NO:6.

The present invention especially relates to a HIV-1 Pol mutant such as IA-Pol (SEQ ID NO:4) which comprises a leader peptide, such as the human tPA leader, at the amino terminal portion of the protein, which may effect cellular trafficking and hence, immunogenicity of the expressed protein within the host cell. Any such HIV-1 DNA pol mutant disclosed in the above paragraphs is suitable for fusion downstream of a leader peptide, including but by no means limited to the human tPA leader sequence. Therefore, any such leader peptide-based HIV-1 pol mutant construct may include but is not limited to a mutated DNA molecule which effectively alters the catalytic activity of the RT, RNase and/or IN region of the expressed protein, resulting in at least substantially decreased enzymatic activity one or more of the RT, RNase H and/or IN functions of HIV-1 Pol. In a preferred embodiment of this portion of the invention, a leader 20 peptide/HIV-1 DNA pol construct contains a mutation or mutations within the Pol coding region which effectively abolishes RT, RNase H and IN activity. An especially preferable HIV-1 DNA pol construct is a DNA molecule which contains at least one point mutation which alters the active site and catalytic activity within the RT, RNase H and IN domains of Pol, such that each activity is at least substantially abolished, and preferably 25 totally abolished. Such a HIV-1 Pol mutant will most likely comprise at least one point mutation in or around each catalytic domain responsible for RT, RNase H and IN activity, respectfully. An especially preferred embodiment of this portion of the invention relates to a human tPA leader fused to the IA-Pol protein comprising the nine mutations shown in Table 1. The DNA molecule is disclosed herein as SEQ ID NO:7 and the expressed 30 tPA-IA Pol protein comprises a fusion junction as shown in Figure 3. The complete amino acid sequence of the expressed protein is set forth in SEQ ID NO:8.

The present invention also relates to a substantially purified protein expressed from the DNA polynucleotide vaccines of the present invention, especially the purified

proteins set forth below as SEQ ID NOs: 2, 4, 6, and 8. These purified proteins may be useful as protein-based HIV vaccines.

The present invention also relates to non-codon optimized versions of DNA molecules and associated polynucleotides and associated DNA vaccines which encode the various wild type and modified forms of the HIV Pol protein disclosed herein. Partial or fully codon optimized DNA vaccine expression vector constructs are preferred, but it is within the scope of the present invention to utilize "non-codon optimized" versions of the constructs disclosed herein, especially modified versions of HIV Pol which are shown to promote a substantial cellular immune and humoral immune responses subsequent to host administration.

5

10

20

25

30

The DNA backbone of the DNA vaccines of the present invention are preferably DNA plasmid expression vectors. DNA plasmid expression vectors utilized in the present invention include but are not limited to constructs which comprise the cytomegalovirus promoter with the intron A sequence (CMV-intA) and 15 a bovine growth hormone transcription termination sequence. In addition, DNA plasmid vectors of the present invention preferably comprise an antibiotic resistance marker, including but not limited to an ampicillin resistance gene, a neomycin resistance gene or any other pharmaceutically acceptable antibiotic resistance marker. In addition, an appropriate polylinker cloning site and a prokaryotic origin of replication sequence are also preferred. Specific DNA vectors exemplified herein include V1, V1J (SEQ ID NO:13), V1Jneo (SEQ ID NO:14), V1Jns (Figure 1A, SEO ID NO:15), V1R (SEQ ID NO:26), and any of the aforementioned vectors wherein a nucleotide sequence encoding a leader peptide, preferably the human tPA leader, is fused directly downstream of the CMV-intA promoter, including but not limited to V1Jns-tpa, as shown in Figure 1B and SEQ ID NO:28.

The present invention especially relates to a DNA vaccine and a pharmaceutically active vaccine composition which contains this DNA vaccine, and the use as prophylactic and/or therapeutic vaccine for host immunization, preferably human host immunization, against an HIV infection or to combat an existing HIV condition. These DNA vaccines are represented by codon optimized DNA molecules encoding codon optimized HIV-1 Pol (e.g. SEQ ID NO:2), codon optimized HIV-1 Pol fused to an amino terminal localized leader sequence (e.g. SEQ ID NO:6), and especially preferable, and the essence of the present invention, biologically inactive Pol proteins (IA Pol; e.g., SEQ ID NO:4) devoid of significant PR, RT, RNase or IN

activity associated with wild type Pol and a concomitant construct which contains a leader peptide at the amino terminal region of the IA Pol protein. These constructs are ligated within an appropriate DNA plasmid vector, with or without a nucleotide sequence encoding a functional leader peptide. Preferred DNA vaccines of the present invention comprise codon optimized DNA molecules encoding codon optimized HIV-1 Pol and inactivated version of Pol, ligated in DNA vectors disclosed herein, or any of the aforementioned vectors wherein a nucleotide sequence encoding a leader peptide, preferably the human tPA leader, is fused directly downstream of the CMV-intA promoter, including but not limited to V1Jns-tpa, as shown in Figure 1B and SEQ ID NO:28.

5

10

15

: 20

25

30

Therefore, the present invention relates to DNA vaccines which include, but are in no way limited to V1Jns-WTPol (comprising the DNA molecule encoding WT Pol, as set forth in SEQ ID NO:2), V1Jns-tPA-WTPol, (comprising the DNA molecule encoding tPA Pol, as set forth in SEQ ID NO:6), V1Jns-IAPol (comprising the DNA molecule encoding IA Pol, as set forth in SEQ ID NO:4), and V1Jns-tPA-IAPol, (comprising the DNA molecule encoding tPA-IA Pol, as set forth in SEQ ID NO:8). Especially preferred are V1Jns-IAPol and V1Jns-tPA-IAPol, as exemplified in Example Section 2.

The present invention also relates to HIV Pol polynucleotide pharmaceutical products, as well as the production and use thereof, wherein the DNA vaccines are formulated with an adjuvant or adjuvants which may increase immunogenicity of the DNA polynucleotide vaccines of the present invention, namely by promoting an enhanced cellular and/or humoral response subsequent to inoculation. A preferred adjuvant is an aluminum phosphate-based adjuvant or a calcium phosphate based adjuvant, with an aluminum phosphate adjuvant being especially preferred. Another preferred adjuvant is a non-ionic block copolymer, preferably comprising the blocks of polyoxyethylene (POE) and polyoxypropylene (POP) such as a POE-POP-POE block copolymer. These adjuvanted forms comprising the DNA vaccines disclosed herein are useful in increasing cellular responses to DNA vaccination.

As used herein, a DNA vaccine or DNA polynucleotide vaccine is a DNA molecule (i.e., "nucleic acid", "polynucleotide") which contains essential regulatory elements such that upon introduction into a living, vertebrate cell, it is able to direct the cellular machinery to produce translation products encoded by the respective pol

genes of the present invention.

5

10

15

20

25

30

BRIEF DESCRIPTION OF THE FIGURES

Figure 1A-B shows schematic representation of DNA vaccine expression vectors V1Jns (A) and V1Jns-tPA (B) utilized for HIV-1 pol and HIV-1 modified pol constructs.

Figure 2A-C shows the nucleotide (SEQ ID NO:3) and amino acid sequence (SEQ ID NO:4) of IA-Pol. Underlined codons and amino acids denote mutations, as listed in Table 1.

Figure 3 shows the codon optimized nucleotide and amino acid sequences through the fusion junction of tPA-IA-Pol (contained within SEQ ID NOs: 7 and 8, respectively). The underlined portion represents the NH₂-terminal region of IA-Pol.

Figure 4 shows generation of a humoral response (measured as the geometric means of anti-RT endpoint titers) from mice immunized with one or two doses of codon optimized V1Jns-IApol and V1Jns-tpa-IApol. A portion of mice that received 30 ug of each plasmid was boosted at T=8 wks; sera from all mice were collected at 4 wk post dose 2.

Figure 5 shows the number of IFN-gamma secreting cells per 10e6 cells following stimulation with pools of either CD4⁺ (aa641-660, aa731-750) or CD8⁺ (aa201-220, aa311-330, aa571-590, aa781-800) specific peptides of splenocytes (pool of 5 spleens/cohort) from control mice and those vaccinated with increasing single dose of codon optimized V1Jns-IApol or 30 ug of codon optimized V1Jns-tpa-IApol (13 wks post dose 1). Mice (n=5) vaccinated with a second dose of 30 ug of either plasmid were analyzed in an Elispot assay at 6 wks post dose 2. Reported are the sums of the number of spots stimulated by each individual CD8⁺ peptides because the spots in the wells to which the pool was added are too dense to acquire accurate counts. The CD4⁺ cell counts are taken from the responses to the peptide pool. Error bars represent standard deviations for counts from triplicate wells per sample per antigen.

Figure 6A-C shows ELIspot analysis of peripheral blood cells collected from rhesus macaques immunized three times (T=0, 4, 8 wks) with 5 mgs of codon optimized HIV-1 Pol expressing plasmids. Antigen-specific IFN-gamma secretion was stimulated by adding one of two pools consisting of 20-mer peptides derived from vaccine sequence (mpol-1, aa1-420; mpol-2, aa411-850). (A) Frequencies of

spot-forming cells (SFC) as a function of time for 3 monkeys (Tag No. 94R008, 94R013, 94R033) vaccinated with V1Jns-IApol. The reported values are corrected for background responses without peptide restimulation. (B) Frequencies of spot-forming cells (SFC) as a function of time for 3 monkeys (Tag No. 920078, 920073, 94R028) vaccinated with 5mgs of V1Jns-tpa-IApol. (C) ELIspot responses were also measured from a monkey (920072) that did not receive any immunization.

5

10

15

20

25

30

Figure 7A-B show bulk CTL killing from rhesus macaques immunized with codon optimized V1Jns-IApol (A)or codon optimized V1Jns-tpa-IApol (B) at 8 weeks following the third vaccination. Restimulation was performed using recombinant vaccinia virus expressing pol and target cells were prepared by pulsing with the peptide pools, mpol-1 and mpol-2.

Figure 8 shows detection of *in vitro* pol expression from cell lysates of 293 cells transfected with 10 ug of various pol constructs. Bands were detected using antiserum from an HIV-1 seropositive human subject. Equal amounts of total protein were loaded for each lane. The lanes contain the lysates from cells transfected with the following: 1: mock; 2: V1Jns-wt-pol; 3: V1Jns-IApol (codon optimized); 4: V1Jns-tpa-IApol (codon optimized); 5: V1Jns-tpa-pol (codon optimized); 6: V1R-wt-pol (codon optimized); 7: blank; and 8: 80 ng RT.

Figure 9 shows the geometric mean anti-RT titers (GMT) plus the standard errors of the geometric means for cohorts of 5 mice that received one (open circles) or two doses (solid circles) of 1, 10, 100 µg of V1R-wt-pol (codon optimized) or V1Ins-wt-pol. Sera from all animals were collected at 2 weeks post dose 2 (or 7 wks post dose 1) and assayed simultaneously. Statistical analyses were performed to compare cohorts that received the same amount and number of immunization of either plasmids; p values (two-tail) less than 5% are above the bars the connect the correlated cohorts to reflect statistically significant differences.

Figure 10 shows cellular immune responses in BALB/c mice vaccinated i.m. with 1 (pd1) or 2 (pd2) doses of varying amounts of either wt-pol (virus derived) or wt-pol (codon optimized) plasmids. At 3 wks post dose 2, frequencies of IFN- γ -secreting splenocytes are determined from pools of 5 spleens per cohort against mixtures of either CD4⁺ peptides (aa21-40, aa411-430, aa531-550, aa641-660, aa731-750, aa771-790) or CD8⁺ peptides (aa201-220, aa311-330) at 4 μ g/mL final concentration per peptide.

DETAILED DESCRIPTION OF THE INVENTION

20

25

30

The present invention relates to synthetic DNA molecules and associated DNA vaccines which elicit CTL and Th cellular immune responses upon administration to the host, including primates and especially humans. An effect of the 5 cellular immune-directed vaccines of the present invention should be a lower transmission rate to previously uninfected individuals and/or reduction in the levels of the viral loads within an infected individual, so as to prolong the asymptomatic phase of HIV-1 infection. In particular, the present invention relates to DNA vaccines which encode various forms of HIV-1 Pol, wherein administration, intracellular 10 delivery and expression of the HIV-1 Pol gene of interest elicits a host CTL and Th response. The preferred synthetic DNA molecules of the present invention encode codon optimized wild type Pol (without Pro activity) and various codon optimized inactivated HIV-1 Pol proteins. The HIV-1 pol constructs disclosed herein are especially preferred for pharmaceutical uses, especially for human administration as a 15 DNA vaccine. The HIV-1 genome employs predominantly uncommon codons compared to highly expressed human genes. Therefore, the pol open reading frame has been synthetically manipulated using optimal codons for human expression. As noted above, a preferred embodiment of the present invention relates to DNA molecules which comprise a HIV-1 pol open reading frame, whether encoding full length pol or a modification or fusion as described herein, wherein the codon usage has been optimized for expression in a mammal, especially a human.

The synthetic pol gene disclosed herein comprises the coding sequences for the reverse transcriptase (or RT which consists of a polymerase and RNase H activity) and integrase (IN). The protein sequence is based on that of Hxb2r, a clonal isolate of IIIB; this sequence has been shown to be closest to the consensus clade B sequence with only 16 nonidentical residues out of 848 (Korber, et al., 1998, Human retroviruses and AIDS, Los Alamos National Laboratory, Los Alamos, New Mexico). The skilled artisan will understand after review of this specification that any available HIV-1 or HIV-2 strain provides a potential template for the generation of HIV pol DNA vaccine constructs disclosed herein. It is further noted that the protease gene is excluded from the DNA vaccine constructs of the present invention to insure safety from any residual protease activity in spite of mutational inactivation. The design of the gene sequences for both wild-type (wt-pol) and inactivated pol (IA-pol) incorporates the use of human preferred ("humanized") codons for each amino acid

residue in the sequence in order to maximize in vivo mammalian expression (Lathe, 1985, J. Mol. Biol. 183:1-12). As can be discerned by inspecting the codon usage in SEQ ID NOs: 1, 3, 5 and 7, the following codon usage for mammalian optimization is preferred: Met (ATG), Gly (GGC), Lys (AAG), Trp (TGG), Ser (TCC), Arg (AGG), Val (GTG), Pro (CCC), Thr (ACC), Glu (GAG); Leu (CTG), His (CAC), Ile (ATC), 5 Asn (AAC), Cys (TGC), Ala (GCC), Gln (CAG), Phe (TTC) and Tyr (TAC). For an additional discussion relating to mammalian (human) codon optimization, see WO 97/31115 (PCT/US97/02294), which is hereby incorporated by reference. It is intended that the skilled artisan may use alternative versions of codon optimization or may omit this step when generating HIV pol vaccine constructs within the scope of 10 the present invention. Therefore, the present invention also relates to non-codon optimized versions of DNA molecules and associated DNA vaccines which encode the various wild type and modified forms of the HIV Pol protein disclosed herein. However, codon optimization of these constructs is a preferred embodiment of this invention. 15

A particular embodiment of the present invention relates to codon optimized wt-pol DNA constructs (herein, "wt-pol" or "wt-pol (codon optimized))" wherein DNA sequences encoding the protease (PR) activity are deleted, leaving codon optimized "wild type" sequences which encode RT (reverse transcriptase and RNase 20 H activity) and IN integrase activity. A DNA molecule which encodes this protein is disclosed herein as SEQ ID NO:1, the open reading frame being contained from an initiating Met residue at nucleotides 10-12 to a termination codon from nucleotides 2560-2562. SEQ ID NO:1 is as follows: AGATCTACCA TGGCCCCCAT CTCCCCCATT GAGACTGTGC CTGTGAAGCT GAAGCCTGGC ATGGATGGCC CCAAGGTGAA GCAGTGGCCC CTGACTGAGG AGAAGATCAA GGCCCTGGTG 25 GAAATCTGCA CTGAGATGGA GAAGGAGGGC AAAATCTCCA AGATTGGCCC CGAGAACCCC TACAACACCC CTGTGTTTGC CATCAAGAAG AAGGACTCCA CCAAGTGGAG GAAGCTGGTG GACTTCAGGG AGCTGAACAA GAGGACCCAG GACTTCTGGG AGGTGCAGCT GGGCATCCCC CACCCCGCTG GCCTGAAGAA GAAGAAGTCT GTGACTGTGC TGGATGTGGG GGATGCCTAC TTCTCTGTGC CCCTGGATGA GGACTTCAGG AAGTACACTG CCTTCACCAT CCCCTCCATC 30 AACAATGAGA CCCCTGGCAT CAGGTACCAG TACAATGTGC TGCCCCAGGG CTGGAAGGGC TCCCCTGCCA TCTTCCAGTC CTCCATGACC AAGATCCTGG AGCCCTTCAG GAAGCAGAAC

CCTGACATTG TGATCTACCA GTACATGGAT GACCTGTATG TGGGCTCTGA CCTGGAGATT
GGGCAGCACA GGACCAAGAT TGAGGAGCTG AGGCAGCACC TGCTGAGGTG GGGCCTGACC

- ACCCCTGACA AGAAGCACCA GAAGGAGCCC CCCTTCCTGT GGATGGGCTA TGAGCTGCAC CCCGACAGT GGACTGTGCA GCCCATTGTG CTGCCTGAGA AGGACTCCTG GACTGTGAAT GACATCCAGA AGCTGGTGGG CAAGCTGAAC TGGGCCTCCC AAATCTACCC TGGCATCAAG GTGAGGCAGC TGTGCAAGCT GCTGAGGGGC ACCAAGGCCC TGACTGAGGT GATCCCCCTG ACTGAGGAGG CTGAGCTGGA GCTGGCTGAG AACAGGGAGA TCCTGAAGGA GCCTGTGCAT GGGGTGTACT ATGACCCCTC CAAGGACCTG ATTGCTGAGA TCCAGAAGCA GGGCCAGGGC CAGTGGACCT ACCAAATCTA CCAGGAGCCC TTCAAGAACC TGAAGACTGG CAAGTATGCC AGGATGAGGG GGGCCCACAC CAATGATGTG AAGCAGCTGA CTGAGGCTGT GCAGAAGATC ACCACTGAGT CCATTGTGAT CTGGGGCAAG ACCCCCAAGT TCAAGCTGCC CATCCAGAAG 10 GAGACCTGGG AGACCTGGTG GACTGAGTAC TGGCAGGCCA CCTGGATCCC TGAGTGGGAG TTTGTGAACA CCCCCCCCT GGTGAAGCTG TGGTACCAGC TGGAGAAGGA GCCCATTGTG GGGGCTGAGA CCTTCTATGT GGATGGGGCT GCCAACAGGG AGACCAAGCT GGGCAAGGCT. GGCTATGTGA CCAACAGGGG CAGGCAGAAG GTGGTGACCC TGACTGACAC CACCAACCAG AAGACTGAGC TCCAGGCCAT CTACCTGGCC CTCCAGGACT CTGGCCTGGA GGTGAACATT GTGACTGACT CCCAGTATGC CCTGGGCATC ATCCAGGCCC AGCCTGATCA GTCTGAGTCT GTGCCTGCCC ACAAGGGCAT TGGGGGCAAT GAGCAGGTGG ACAAGCTGGT GTCTGCTGGC ATCAGGAAGG TGCTGTTCCT GGATGGCATT GACAAGGCCC AGGATGAGCA TGAGAAGTAC CACTCCAACT GGAGGGCTAT GGCCTCTGAC TTCAACCTGC CCCCTGTGGT GGCTAAGGAG 20 ATTGTGGCCT CCTGTGACAA GTGCCAGCTG AAGGGGGAGG CCATGCATGG GCAGGTGGAC TGCTCCCTG GCATCTGGCA GCTGGACTGC ACCCACCTGG AGGGCAAGGT GATCCTGGTG GCTGTGCATG TGGCCTCCGG CTACATTGAG GCTGAGGTGA TCCCTGCTGA GACAGGCCAG GAGACTGCCT ACTTCCTGCT GAAGCTGGCT GGCAGGTGGC CTGTGAAGAC CATCCACACT GACAATGGCT CCAACTTCAC TGGGGCCACA GTGAGGGCTG CCTGCTGGTG GGCTGGCATC 25 AAGCAGGAGT TTGGCATCCC CTACAACCCC CAGTCCCAGG GGGTGGTGGA GTCCATGAAC AAGGAGCTGA AGAAGATCAT TGGGCAGGTG AGGGACCAGG CTGAGCACCT GAAGACAGCT GTGCAGATGG CTGTGTTCAT CCACAACTTC AAGAGGAAGG GGGGCATCGG GGGCTACTCC GCTGGGGAGA GGATTGTGGA CATCATTGCC ACAGACATCC AGACCAAGGA GCTCCAGAAG CAGATCACCA AGATCCAGAA CTTCAGGGTG TACTACAGGG ACTCCAGGAA CCCCCTGTGG 30 AAGGGCCCTG CCAAGCTGCT GTGGAAGGGG GAGGGGGCTG TGGTGATCCA GGACAACTCT GACATCAAGG TGGTGCCCAG GAGGAAGGCC AAGATCATCA GGGACTATGG CAAGCAGATG GCTGGGGATG ACTGTGTGGC CTCCAGGCAG GATGAGGACT AAAGCCCGGG CAGATCT (SEO ID NO:1).

The open reading frame of the wild type pol construct disclosed as SEQ ID NO:1 contains 850 amino acids, disclosed herein as SEQ ID NO:2, as follows: Met Ala Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu. Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys Ser Val Thr Val Leu Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg Lys 10 Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr 20 Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Asp Gly Ala Ala 30 Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Glu Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Asp Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro

Asp Gln Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Asp Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr Ile His Thr Asp Asn Gly Ser Asn Phe Thr Gly Ala Thr Val Arg Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly Val Val Glu Ser Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr 15 Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro Ala Lys Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn 20 Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp (SEQ ID NO:2).

The present invention especially relates to a codon optimized HIV-1 DNA pol construct wherein, in addition to deletion of the portion of the wild type sequence encoding the protease activity, a combination of active site residue mutations are introduced which are deleterious to HIV-1 pol (RT-RH-IN) activity of the expressed protein. Therefore, the present invention preferably relates to a HIV-1 DNA pol construct which is devoid of DNA sequences encoding any PR activity, as well as containing a mutation(s) which at least partially, and preferably substantially, abolishes RT, RNase and/or IN activity. One type of HIV-1 pol mutant may include but is not limited to a mutated DNA molecule comprising at least one nucleotide substitution which results in a point mutation which effectively alters an active site within the RT, RNase and/or IN regions of the expressed protein, resulting in at least substantially decreased enzymatic activity for the RT, RNase H and/or IN functions of

25

30

5

10

15

20

HIV-1 Pol. In a preferred embodiment of this portion of the invention, a HIV-1 DNA pol construct contains a mutation or mutations within the Pol coding region which effectively abolishes RT, RNase H and IN activity. An especially preferable HIV-1 DNA pol construct in a DNA molecule which contains at least one point mutation which alters the active site of the RT, RNase H and IN domains of Pol, such that each activity is at least substantially abolished. Such a HIV-1 Pol mutant will most likely comprise at least one point mutation in or around each catalytic domain responsible for RT, RNase H and IN activity, respectfully. To this end, an especially preferred HIV-1 DNA pol construct is exemplified herein and contains nine codon substitution mutations which results in an inactivated Pol protein (IA Pol: SEQ ID NO:4, Figure 2A-C) which has no PR, RT, RNase or IN activity, wherein three such point mutations reside within each of the RT, RNase and IN catalytic domains. Therefore, an especially preferred exemplification is a DNA molecule which encodes IA-pol, which contains all nine mutations as shown below in Table 1. An additional preferred amino acid residue for substitution is Asp551, localized within the RNase domain of Pol. Any combination of the mutations disclosed herein may suitable and therefore may be utilized as an IA-Pol-based vaccine of the present invention. While addition and deletion mutations are contemplated and within the scope of the invention, the preferred mutation is a point mutation resulting in a substitution of the wild type amino acid with an alternative amino acid residue.

			Table 1	
	wt aa	aa residue	mutant aa	enzyme function
	Asp	112	Ala	RT
25	Asp	187	Ala	RT
	Asp	188	Ala	RT
	Asp	445	Ala	RNase H
	Glu	480	Ala	RNase H
30	Asp	500	Ala	RNase H
	Asp	626	Ala	IN
	Asp	678	Ala	· IN
	Glu	714	Ala	IN

It is preferred that point mutations be incorporated into the IApol mutant vaccines of the present invention so as to lessen the possibility of altering epitopes in and around the active site(s) of HIV-1 Pol.

To this end, SEQ ID NO:3 discloses the nucleotide sequence which codes for a codon optimized pol in addition to the nine mutations shown in Table 1, disclosed as follows, and referred to herein as "IApol":

5

AGATCTACCA TGGCCCCCAT CTCCCCCATT GAGACTGTGC CTGTGAAGCT GAAGCCTGGC ATGGATGGCC CCAAGGTGAA GCAGTGGCCC CTGACTGAGG AGAAGATCAA GGCCCTGGTG GAAATCTGCA CTGAGATGGA GAAGGAGGGC AAAATCTCCA AGATTGGCCC CGAGAACCCC 10 TACAACACCC CTGTGTTTGC CATCAAGAAG AAGGACTCCA CCAAGTGGAG GAAGCTGGTG GACTTCAGGG AGCTGAACAA GAGGACCCAG GACTTCTGGG AGGTGCAGCT GGGCATCCCC CACCCCGCTG GCCTGAAGAA GAAGAAGTCT GTGACTGTGC TGGCTGTGGG GGATGCCTAC TTCTCTGTGC CCCTGGATGA GGACTTCAGG AAGTACACTG CCTTCACCAT CCCCTCCATC AACAATGAGA CCCCTGGCAT CAGGTACCAG TACAATGTGC TGCCCCAGGG CTGGAAGGGC 15 TCCCCTGCCA TCTTCCAGTC CTCCATGACC AAGATCCTGG AGCCCTTCAG GAAGCAGAAC CCTGACATTG TGATCTACCA GTACATGGCT GCCCTGTATG TGGGCTCTGA CCTGGAGATT GGGCAGCACA GGACCAAGAT TGAGGAGCTG AGGCAGCACC TGCTGAGGTG GGGCCTGACC ACCCCTGACA AGAAGCACCA GAAGGAGCCC CCCTTCCTGT GGATGGGCTA TGAGCTGCAC CCCGACAAGT GGACTGTGCA GCCCATTGTG CTGCCTGAGA AGGACTCCTG GACTGTGAAT 20 GACATCCAGA AGCTGGTGGG CAAGCTGAAC TGGGCCTCCC AAATCTACCC TGGCATCAAG GTGAGGCAGC TGTGCAAGCT GCTGAGGGGC ACCAAGGCCC TGACTGAGGT GATCCCCCTG ACTGAGGAGG CTGAGCTGGA GCTGGCTGAG AACAGGGAGA TCCTGAAGGA GCCTGTGCAT GGGGTGTACT ATGACCCCTC CAAGGACCTG ATTGCTGAGA TCCAGAAGCA GGGCCAGGGC CAGTGGACCT ACCAAATCTA CCAGGAGCCC TTCAAGAACC TGAAGACTGG CAAGTATGCC 25 AGGATGAGGG GGGCCCACAC CAATGATGTG AAGCAGCTGA CTGAGGCTGT GCAGAAGATC ACCACTGAGT CCATTGTGAT CTGGGGCAAG ACCCCCAAGT TCAAGCTGCC CATCCAGAAG GAGACCTGGG AGACCTGGTG GACTGAGTAC TGGCAGGCCA CCTGGATCCC TGAGTGGGAG TTTGTGAACA CCCCCCCCT GGTGAAGCTG TGGTACCAGC TGGAGAAGGA GCCCATTGTG GGGGCTGAGA CCTTCTATGT GGCTGGGGCT GCCAACAGGG AGACCAAGCT GGGCAAGGCT 30 GGCTATGTGA CCAACAGGGG CAGGCAGAAG GTGGTGACCC TGACTGACAC CACCAACCAG AAGACTGCCC TCCAGGCCAT CTACCTGGCC CTCCAGGACT CTGGCCTGGA GGTGAACATT GTGACTGCCT CCCAGTATGC CCTGGGCATC ATCCAGGCCC AGCCTGATCA GTCTGAGTCT GTGCCTGCCC ACAAGGGCAT TGGGGGCAAT GAGCAGGTGG ACAAGCTGGT GTCTGCTGGC

5

10

15

20

25

30

ATCAGGAAGG TGCTGTTCCT GGATGGCATT GACAAGGCCC AGGATGAGCA TGAGAAGTAC CACTCCAACT GGAGGGCTAT GGCCTCTGAC TTCAACCTGC CCCCTGTGGT GGCTAAGGAG ATTGTGGCCT CCTGTGACAA GTGCCAGCTG AAGGGGGAGG CCATGCATGG GCAGGTGGAC TGCTCCCTG GCATCTGGCA GCTGGCCTGC ACCCACCTGG AGGGCAAGGT GATCCTGGTG GCTGTGCATG TGGCCTCCGG CTACATTGAG GCTGAGGTGA TCCCTGCTGA GACAGGCCAG GAGACTGCCT ACTTCCTGCT GAAGCTGGCT GGCAGGTGGC CTGTGAAGAC CATCCACACT GCCAATGGCT CCAACTTCAC TGGGGCCACA GTGAGGGCTG CCTGCTGGTG GGCTGGCATC AAGCAGGAGT TTGGCATCCC CTACAACCCC CAGTCCCAGG GGGTGGTGGC CTCCATGAAC AAGGAGCTGA AGAAGATCAT TGGGCAGGTG AGGGACCAGG CTGAGCACCT GAAGACAGCT GTGCAGATGG CTGTGTTCAT CCACAACTTC AAGAGGAAGG GGGGCATCGG GGGCTACTCC GCTGGGGAGA GGATTGTGGA CATCATTGCC ACAGACATCC AGACCAAGGA GCTCCAGAAG CAGATCACCA AGATCCAGAA CTTCAGGGTG TACTACAGGG ACTCCAGGAA CCCCCTGTGG AAGGGCCCTG CCAAGCTGCT GTGGAAGGGG GAGGGGGCTG TGGTGATCCA GGACAACTCT GACATCAAGG TGGTGCCCAG GAGGAAGGCC AAGATCATCA GGGACTATGG CAAGCAGATG GCTGGGGATG ACTGTGTGGC CTCCAGGCAG GATGAGGACT AAAGCCCGGG CAGATCT (SEQ ID NO:3).

In order to produce the IA-pol DNA vaccine construction, inactivation of the enzymatic functions was achieved by replacing a total of nine active-site residues from the enzyme subunits with alanine side-chains. As shown in Table 1, all residues that comprise the catalytic triad of the polymerase, namely Asp112, Asp187, and Asp188, were substituted with alanine (Ala) residues (Larder, et al., Nature 1987, 327: 716-717; Larder, et al., 1989, Proc. Natl. Acad. Sci. 1989, 86: 4803-4807). Three additional mutations were introduced at Asp445, Glu480 and Asp500 to abolish RNase H activity (Asp551 was left unchanged in this IA Pol construct), with each residue being substituted for an Ala residue, respectively (Davies, et al., 1991, Science 252:, 88-95; Schatz, et al., 1989, FEBS Lett. 257: 311-314; Mizrahi, et al., 1990, Nucl. Acids. Res. 18: pp. 5359-5353). HIV pol integrase function was abolished through three mutations at Asp626, Asp678 and Glu714. Again, each of these residues has been substituted with an Ala residue (Wiskerchen, et al., 1995, J. Virol. 69: 376-386; Leavitt, et al., 1993, J. Biol. Chem. 268: 2113-2119). Amino acid residue Pro3 of SEQ ID NO:4 marks the start of the RT gene. The complete amino acid sequence of IA-Pol is disclosed herein as SEQ ID NO:4, as follows: Met Ala Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys

Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys Ser Val Thr Val Leu Ala Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln 10 Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Ala Ala Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val 15 Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln 20 Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp 25 Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Ala Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Ala Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn 30 Ile Val Thr Ala Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Gln Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys

Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Ala Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly 5 Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr Ile His Thr Ala Asn Gly Ser Asn Phe Thr Gly Ala Thr Val Arg Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly Val Val Ala Ser Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr 10 Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn 15 Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp (SEQ ID NO:4).

As noted above, it will be understood that any combination of the mutations disclosed above may be suitable and therefore be utilized as an IA-pol-based vaccine of the present invention. For example, it may be possible to mutate only 2 of the 3 residues within the respective reverse transcriptase, RNase H, and integrase coding regions while still abolishing these enzymatic activities. However, the IA-pol construct described above and disclosed as SEQ ID NO:3, as well as the expressed protein (SEQ ID NO:4) is preferred. It is also preferred that at least one mutation be present in each of the three catalytic domains.

20

25

30

Another aspect of the present invention is to generate codon optimized HIV-1 Pol-based vaccine constructions which comprise a eukaryotic trafficking signal peptide such as from tPA (tissue-type plasminogen activator) or by a leader peptide such as is found in highly expressed mammalian proteins such as immunoglobulin leader peptides. Any functional leader peptide may be tested for efficacy. However, a preferred embodiment of the present invention is to provide for HIV-1 Pol mutant vaccine constructions as disclosed herein which also comprise a leader peptide, preferably a leader peptide from human tPA. In other words, a codon optimized

HIV-1 Pol mutant such as IA-Pol (SEQ ID NO:4) may also comprise a leader peptide at the amino terminal portion of the protein, which may effect cellular trafficking and hence, immunogenicity of the expressed protein within the host cell. As shown in Figure 1A-B for the DNA vector V1Jns, a DNA vector which may be utilized to practice the present invention may be modified by known recombinant DNA methodology to contain a leader signal peptide of interest, such that downstream cloning of the modified HIV-1 protein of interest results in a nucleotide sequence which encodes a modified HIV-1 tPA/Pol protein. In the alternative, as noted above. insertion of a nucleotide sequence which encodes a leader peptide may be inserted into a DNA vector housing the open reading frame for the Pol protein of interest. Regardless of the cloning strategy, the end result is a polynucleotide vaccine which comprises vector components for effective gene expression in conjunction with nucleotide sequences which encode a modified HIV-1 Pol protein of interest, including but not limited to a HIV-1 Pol protein which contains a leader peptide. The amino acid sequence of the human tPA leader utilized herein is as follows: MDAMKRGLCCVLLLCGAVFVSPSEISS (SEQ ID NO:28). Therefore, another aspect of the present invention is to generate HIV-1 Pol-based vaccine constructions which comprise a eukaryotic trafficking signal peptide such as from tPA. To this end, the present invention relates to a DNA molecule which encodes a codon optimized wt-pol DNA construct wherein the protease (PR) activity is deleted and a human tPA leader sequence is fused to the 5' end of the coding region. A DNA molecule which encodes this protein is disclosed herein as SEQ ID NO:5, the open reading frame disclosed herein as SEQ ID NO:6.

To this end, the present invention relates to a DNA molecule which encodes a codon optimized wt-pol DNA construct wherein the protease (PR) activity is deleted and a human tPA leader sequence is fused to the 5' end of the coding region (herein, "tPA-wt-pol"). A DNA molecule which encodes this protein is disclosed herein as SEQ ID NO:5, the open reading frame being contained from an initiating Met residue at nucleotides 8-10 to a termination codon from nucleotides 2633-2635. SEQ ID

30 NO:5 is as follows:

5

10

15

20

25

GATCACCATG GATGCAATGA AGAGAGGGCT CTGCTGTGT CTGCTGTGT GTGGAGCAGT
CTTCGTTTCG CCCAGCGAGA TCTCCGCCCC CATCTCCCCC ATTGAGACTG TGCCTGTGAA
GCTGAAGCCT GGCATGGATG GCCCCAAGGT GAAGCAGTGG CCCCTGACTG AGGAGAAGAT
CAAGGCCCTG GTGGAAATCT GCACTGAGAT GGAGAAGGAG GGCAAAATCT CCAAGATTGG

	CCCCGAGAAC					
				CAAGAGGACC		
				GAAGAAGAAG		
	GGGGGATGCC	TACTTCTCTG	TGCCCCTGGA	TGAGGACTTC	AGGAAGTACA	CTGCCTTCAC
5	CATCCCCTCC	ATCAACAATG	AGACCCCTGG	CATCAGGTAC	CAGTACAATG	TGCTGCCCCA
	GGGCTGGAAG	GGCTCCCCTG	CCATCTTCCA	GTCCTCCATG	ACCAAGATCC	TGGAGCCCTT
	CAGGAAGCAG	AACCCTGACA	TTGTGATCTA	CCAGTACATG	GATGACCTGT	ATGTGGGCTC
	TGACCTGGAG	ATTGGGCAGC	ACAGGACCAA	GATTGAGGAG	CTGAGGCAGC	ACCTGCTGAG
	GTGGGGCCTG	ACCACCCCTG	ACAAGAAGCA	CCAGAAGGAG	CCCCCCTTCC	TGTGGATGGG
10	CTATGAGCTG	CACCCGACA	AGTGGACTGT	GCAGCCCATT	GTGCTGCCTG	AGAAGGACTC
	CTGGACTGTG	AATGACATCC	AGAAGCTGGT	GGGCAAGCTG	AACTGGGCCT	CCCAAATCTA
	CCCTGGCATC	AAGGTGAGGC	AGCTGTGCAA	GCTGCTGAGG	GGCACCAAGG	CCCTGACTGA
				GGAGCTGGCT		
				CTCCAAGGAC		
15				CTACCAGGAG		
				CACCAATGAT		
				GATCTGGGGC		
				GTGGACTGAG		
				CCTGGTGAAG		
20				TGTGGATGGG		
				GGGCAGGCAG		
				CATCTACCTG		
				TGCCCTGGGC		
				CATTGAGCAG		
25				CATTGGGGGC		
				r CCTGGATGGC		
						TGCCCCCTGT
				A CAAGTGCCAG		
						TGGAGGGCAA
30						TGATCCCTGC
						r GGCCTGTGAA
						G CTGCCTGCTG
						C AGGGGGTGGT
	GGAGTCCAT	G AACAAGGAG	C TGAAGAAGA	T CATTGGGCA	G GTGAGGGAC	C AGGCTGAGCA

CCTGAAGACA GCTGTGCAGA TGGCTGTGTT CATCCACAAC TTCAAGAGGA AGGGGGGCAT
CGGGGGCTAC TCCGCTGGGG AGAGGATTGT GGACATCATT GCCACAGACA TCCAGACCAA
GGAGCTCCAG AAGCAGATCA CCAAGATCCA GAACTTCAGG GTGTACTACA GGGACTCCAG
GAACCCCCTG TGGAAGGGCC CTGCCAAGCT GCTGTGGAAG GGGGAGGGGG CTGTGGTGAT
CCAGGACAAC TCTGACATCA AGGTGGTGCC CAGGAGGAAG GCCAAGATCA TCAGGGACTA
TGGCAAGCAG ATGGCTGGGG ATGACTGTGT GGCCTCCAGG CAGGATGAGG ACTAAAGCCC
GGGCAGATCT (SEQ ID NO:5).

The open reading frame of the wild type tPA-pol construct disclosed as SEQ ID NO:5 contains 875 amino acids, disclosed herein as SEQ ID NO:6, as follows:

10 Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Cys Gly Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ala Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu 15 Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys Ser Val Thr Val Leu Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro 20 Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly 25. Leu Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile 30 Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln

Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Asp Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly 5 Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Glu Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Asp Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Gln Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu 10 Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln 15 Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Asp Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr Ile His Thr Asp Asn Gly Ser Asn Phe Thr Gly Ala Thr Val Arg Ala Ala Cys Trp Trp Ala 20 Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly Val Val Glu Ser Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu 25 Gln Lys Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp (SEQ ID NO:6). 30

The present invention also relates to a codon optimized HIV-1 Pol mutant such as IA-Pol (SEQ ID NO:4) which comprises a leader peptide at the amino terminal portion of the protein, which may effect cellular trafficking and hence, immunogenicity of the expressed protein within the host cell. Any such HIV-1 DNA

pol mutant disclosed in the above paragraphs is suitable for fusion downstream of a leader peptide, such as a leader peptide including but not limited to the human tPA leader sequence. Therefore, any such leader peptide-based HIV-1 pol mutant construct may include but is not limited to a mutated DNA molecule which effectively alters the catalytic activity of the RT, RNase and/or IN region of the expressed protein, resulting in at least substantially decreased enzymatic activity one or more of the RT, RNase H and/or IN functions of HIV-1 Pol. In a preferred embodiment of this portion of the invention, a leader peptide/HIV-1 DNA pol construct contains a mutation or mutations within the Pol coding region which effectively abolishes RT, RNase H and IN activity. An especially preferable HIV-1 DNA pol construct is a DNA molecule 10 which contains at least one point mutation which alters the active site and catalytic activity within the RT, RNase H and IN domains of Pol, such that each activity is at least substantially abolished, and preferably totally abolished. Such a HIV-1 Pol mutant will most likely comprise at least one point mutation in or around each catalytic domain responsible for RT, RNase H and IN activity, respectfully. An especially preferred embodiment of this portion of the invention relates to a human tPA leader fused to the IA-Pol protein comprising the nine mutations shown in Table 1. The DNA molecule is disclosed herein as SEQ ID NO:7 and the expressed tPA-IA Pol protein comprises a fusion junction as shown in Figure 3. The complete amino 20 acid sequence of the expressed protein is set forth in SEQ ID NO:8. To this end, SEQ ID NO:7 discloses the nucleotide sequence which codes for a human tPA leader fused to the IA Pol protein comprising the nine mutations shown in Table 1 (herein, "tPAopt-IApol"). The open reading frame begins with the initiating Met (nucleotides 8-10) and terminates with a "TAA" codon at nucleotides 2633-2635. The nucleotide 25 sequence encoding tPA-IAPol is also disclosed as follows: GATCACCATG GATGCAATGA AGAGAGGGCT CTGCTGTGTG CTGCTGCTGT GTGGAGCAGT CTTCGTTTCG CCCAGCGAGA TCTCCGCCCC CATCTCCCCC ATTGAGACTG TGCCTGTGAA GCTGAAGCCT GGCATGGATG GCCCCAAGGT GAAGCAGTGG CCCCTGACTG AGGAGAAGAT CAAGGCCCTG GTGGAAATCT GCACTGAGAT GGAGAAGGAG GGCAAAATCT CCAAGATTGG 30 CCCCGAGAAC CCCTACAACA CCCCTGTGTT TGCCATCAAG AAGAAGGACT CCACCAAGTG GAGGAAGCTG GTGGACTTCA GGGAGCTGAA CAAGAGGACC CAGGACTTCT GGGAGGTGCA GCTGGGCATC CCCCACCCCG CTGGCCTGAA GAAGAAGAAG TCTGTGACTG TGCTGGCTGT GGGGGATGCC TACTTCTCTG TGCCCCTGGA TGAGGACTTC AGGAAGTACA CTGCCTTCAC CATCCCCTCC ATCAACAATG AGACCCCTGG CATCAGGTAC CAGTACAATG TGCTGCCCCA

	GGGCTGGAAG GGCTCCCCTG	CCATCTTCCA	GTCCTCCATG	ACCAAGATCC '	TGGAGCCCTT
	CAGGAAGCAG AACCCTGACA	TTGTGATCTA	CCAGTACATG	GCTGCCCTGT	ATGTGGGCTC
	TGACCTGGAG ATTGGGCAGC				
	GTGGGGCCTG ACCACCCCTG				
5	CTATGAGCTG CACCCCGACA	AGTGGACTGT	GCAGCCCATT	GTGCTGCCTG	AGAAGGACTC
	CTGGACTGTG AATGACATCC	AGAAGCTGGT	GGGCAAGCTG	AACTGGGCCT	CCCAAATCTA
	CCCTGGCATC AAGGTGAGGC	AGCTGTGCAA	GCTGCTGAGG	GGCACCAAGG	CCCTGACTGA
	GGTGATCCCC CTGACTGAGG	AGGCTGAGCT	GGAGCTGGCT	GAGAACAGGG	AGATCCTGAA
	GGAGCCTGTG CATGGGGTGT	ACTATGACCC	CTCCAAGGAC	CTGATTGCTG	AGATCCAGAA
10	GCAGGGCCAG GGCCAGTGG				
	TGGCAAGTAT GCCAGGATGA	GGGGGCCCA	CACCAATGAT	GTGAAGCAGC	TGACTGAGGC
	TGTGCAGAAG ATCACCACTO	G AGTCCATTGT	GATCTGGGGC	AAGACCCCCA	AGTTCAAGCT
	GCCCATCCAG AAGGAGACC				
	CCCTGAGTGG GAGTTTGTG	A ACACCCCCC	CCTGGTGAAG	CTGTGGTACC	AGCTGGAGAA
15	GGAGCCCATT GTGGGGGCT				
	GCTGGGCAAG GCTGGCTAT				
	CACCACCAAC CAGAAGACT	G CCCTCCAGGC	CATCTACCTG	GCCCTCCAGG	ACTCTGGCCT
	GGAGGTGAAC ATTGTGACT				
	TCAGTCTGAG TCTGAGCTG				
20	GTACCTGGCC TGGGTGCCT				
	GGTGTCTGCT GGCATCAGG				
	GCATGAGAAG TACCACTCC	A ACTGGAGGGC	TATGGCCTCT	GACTTCAACC	TGCCCCCTGT
	GGTGGCTAAG GAGATTGTG				
	TGGGCAGGTG GACTGCTCC				
25	GGTGATCCTG GTGGCTGTG				
•	TGAGACAGGC CAGGAGACT	G CCTACTTCCT	GCTGAAGCTG	GCTGGCAGGT	GGCCTGTGAA
	GACCATCCAC ACTGCCAAT	G GCTCCAACT	CACTGGGGCC	ACAGTGAGGG	CTGCCTGCTG
	GTGGGCTGGC ATCAAGCAC	G AGTTTGGCAT	CCCCTACAAC	CCCCAGTCCC	AGGGGGTGGT
	GGCCTCCATG AACAAGGA	C TGAAGAAGA	r CATTGGGCAC	GTGAGGGACC	AGGCTGAGCA
30	CCTGAAGACA GCTGTGCA				
20	CGGGGGCTAC TCCGCTGG				
	GGAGCTCCAG AAGCAGAT				
	GAACCCCCTG TGGAAGGG	CC CTGCCAAGC	r GCTGTGGAA	GGGGAGGGG	CTGTGGTGAT
	CCAGGACAAC TCTGACAT	CA AGGTGGTGC	C CAGGAGGAA	G GCCAAGATC	A TCAGGGACTA
	0017001101111 = 11.0111				

TGGCAAGCAG ATGGCTGGGG ATGACTGTGT GGCCTCCAGG CAGGATGAGG ACTAAAGCCC GGGCAGATCT (SEQ ID NO:7).

5

The open reading frame of the tPA-IA-pol construct disclosed as SEQ ID NO:7 contains 875 amino acids, disclosed herein as tPA-IA-Pol and SEQ ID NO:8, as follows:

Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Cys Gly Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ala Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val Glu Ile 10 Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu Ala Val Gly Asp Ala Tyr Phe Ser 15 Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Ala Ala Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln 20 His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg 25 Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met 30 Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala

PCT/US00/34724

Glu Thr Phe Tyr Val Ala Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Ala Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Ala Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Gln Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val 10 Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Ala Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr Ile His Thr Ala Asn Gly Ser Asn Phe Thr Gly Ala Thr Val Arg Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly Val Val Ala Ser Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly 20 Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly .25 Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp (SEQ ID NO:8). The present invention also relates to a substantially purified protein expressed

WO 01/45748

30

The present invention also relates to a substantially purified protein expressed from the DNA polynucleotide vaccines of the present invention, especially the purified proteins set forth below as SEQ ID NOs: 2, 4, 6, and 8. These purified proteins may be useful as protein-based HIV vaccines.

The DNA backbone of the DNA vaccines of the present invention are preferably DNA plasmid expression vectors. DNA plasmid expression vectors are well known in the art and the present DNA vector vaccines may be comprised of any such expression backbone which contains at least a promoter for RNA polymerase

transcription, and a transcriptional terminator 3' to the HIV pol coding sequence. In one preferred embodiment, the promoter is the Rous sarcoma virus (RSV) long terminal repeat (LTR) which is a strong transcriptional promoter. A more preferred promoter is the cytomegalovirus promoter with the intron A sequence (CMV-intA). A preferred transcriptional terminator is the bovine growth hormone terminator. In addition, to assist in large scale preparation of an HIV pol DNA vector vaccine, an antibiotic resistance marker is also preferably included in the expression vector. Ampicillin resistance genes, neomycin resistance genes or any other pharmaceutically acceptable antibiotic resistance marker may be used. In a preferred embodiment of 10 this invention, the antibiotic resistance gene encodes a gene product for neomycin resistance. Further, to aid in the high level production of the pharmaceutical by fermentation in prokaryotic organisms, it is advantageous for the vector to contain an origin of replication and be of high copy number. Any of a number of commercially available prokaryotic cloning vectors provide these benefits. In a preferred 15 embodiment of this invention, these functionalities are provided by the commercially available vectors known as pUC. It is desirable to remove non-essential DNA sequences. Thus, the lacZ and lacI coding sequences of pUC are removed in one embodiment of the invention.

DNA expression vectors which exemplify but in no way limit the present invention are disclosed in PCT International Application No. PCT/US94/02751, International Publication No. WO 94/21797, hereby incorporated by reference. A first DNA expression vector is the expression vector pnRSV, wherein the rous sarcoma virus (RSV) long terminal repeat (LTR) is used as the promoter. A second embodiment relates to plasmid V1, a mutated pBR322 vector into which the CMV promoter and the BGH transcriptional terminator is cloned. Another embodiment regarding DNA vector backbones relates to plasmid V1J. Plasmid V1J is derived from plasmid V1 and removes promoter and transcription termination elements in order to place them within a more defined context, create a more compact vector, and to improve plasmid purification yields. Therefore, V1J also contains the CMVintA promoter and (BGH) transcription termination elements which control the expression of the HIV pol-based genes disclosed herein. The backbone of V1J is provided by pUC18. It is known to produce high yields of plasmid, is well-characterized by sequence and function, and is of minimum size. The entire lac operon was removed and the remaining plasmid was purified from an agarose electrophoresis gel,

20

25

30

PCT/US00/34724 WO 01/45748

blunt-ended with the T4 DNA polymerase, treated with calf intestinal alkaline phosphatase, and ligated to the CMVintA/BGH element. In a preferred DNA expression vector, the ampicillin resistance gene is removed from V1J and replaced with a neomycin resistance gene, to generate V1Jneo. An especially preferred DNA expression vector is VIJns, which is the same as VIJ except that a unique Sfi1 5 restriction site has been engineered into the single Kpn1 site at position 2114 of V1Jneo. The incidence of Sfi1 sites in human genomic DNA is very low (approximately 1 site per 100,000 bases). Thus, this vector allows careful monitoring for expression vector integration into host DNA, simply by Sfi1 digestion of extracted genomic DNA. Yet another preferred DNA expression vector used as the backbone to the 10 HIV-1 pol-based DNA vaccines of the present invention is V1R. In this vector, as much non-essential DNA as possible is "trimmed" from the vector to produce a highly compact vector. This vector is a derivative of V1Ins. This vector allows larger inserts to be used, with less concern that undesirable sequences are encoded and optimizes uptake by cells when the construct encoding specific influenza virus genes 15 is introduced into surrounding tissue. The specific DNA vectors of the present invention include but are not limited to V1, V1J (SEQ ID NO:13), V1Jneo (SEQ ID NO:14), V1Jns (Figure 1A, SEQ ID NO:15), V1R (SEQ ID NO:26), and any of the aforementioned vectors wherein a nucleotide sequence encoding a leader peptide, preferably the human tPA leader, is fused directly downstream of the CMV-intA 20 promoter, including but not limited to V1Jns-tpa, as shown in Figure 1B and SEQ ID NO:28.

The present invention especially relates to a DNA vaccine and a pharmaceutically active vaccine composition which contains this DNA vaccine, and the use as prophylactic and/or therapeutic vaccine for host immunization, preferably human host immunization, against an HIV infection or to combat an existing HIV condition. These DNA vaccines are represented by codon optimized DNA molecules encoding HIV-1 Pol or biologically active Pol modifications or Pol-containing fusion proteins which are ligated within an appropriate DNA plasmid vector, with or without a nucleotide sequence encoding a functional leader peptide. DNA vaccines of the present invention may comprise codon optimized DNA molecules encoding HIV-1 Pol or biologically active Pol modifications or Pol-containing fusion proteins ligated in DNA vectors V1, V1J (SEQ ID NO:14), V1Jneo (SEQ ID NO:15), V1Jns (Figure 1A, SEQ ID NO:16), V1R (SEQ ID NO:26), or any of the aforementioned vectors

25

30

wherein a nucleotide sequence encoding a leader peptide, preferably the human tPA leader, is fused directly downstream of the CMV-intA promoter, including but not limited to V1Jns-tpa, as shown in Figure 1B and SEQ ID NO:28. To this end, polynucleotide vaccine constructions include, V1Jns-wtpol and V1R-wtpol (comprising the DNA molecule encoding WT Pol, as set forth in SEQ ID NO:2), V1Jns-tPA-WTPol, (comprising the DNA molecule encoding tPA Pol, as set forth in SEQ ID NO:6), V1Jns-IAPol (comprising the DNA molecule encoding IA Pol, as set forth in SEQ ID NO:4), and V1Jns-tPA-IAPol, (comprising the DNA molecule encoding tPA-IA Pol, as set forth in SEQ ID NO:8). Polynucleotide vaccine constructions V1R-wtpol, V1Jns-IAPol, and V1Jns-tPA-IAPol, are exemplified in Example Sections 3-5.

It will be evident upon review of the teaching within this specification that numerous vector/Pol antigen constructs may be generated. While the exemplified constructs are preferred, any number of vector/Pol antigen combinations are within the scope of the present invention, especially wild type or modified/inactivated Pol proteins which comprise at least one, preferably 5 or more and especially all nine mutations as shown in Table 1, with or without the inclusion of a leader sequence such as human tPA.

15

The DNA vector vaccines of the present invention may be formulated in any 20 pharmaceutically effective formulation for host administration. Any such formulation may be, for example, a saline solution such as phosphate buffered saline (PBS). It will be useful to utilize pharmaceutically acceptable formulations which also provide long-term stability of the DNA vector vaccines of the present invention. During storage as a pharmaceutical entity, DNA plasmid vaccines undergo a 25 physiochemical change in which the supercoiled plasmid converts to the open circular and linear form. A variety of storage conditions (low pH, high temperature, low ionic strength) can accelerate this process. Therefore, the removal and/or chelation of trace metal ions (with succinic or malic acid, or with chelators containing multiple phosphate ligands) from the DNA plasmid solution, from the formulation buffers or 30 from the vials and closures, stabilizes the DNA plasmid from this degradation pathway during storage. In addition, inclusion of non-reducing free radical scavengers, such as ethanol or glycerol, are useful to prevent damage of the DNA plasmid from free radical production that may still occur, even in apparently demetalated solutions. Furthermore, the buffer type, pH, salt concentration, light

5

10

15

25

30

exposure, as well as the type of sterilization process used to prepare the vials, may be controlled in the formulation to optimize the stability of the DNA vaccine. Therefore, formulations that will provide the highest stability of the DNA vaccine will be one that includes a demetalated solution containing a buffer (phosphate or bicarbonate) with a pH in the range of 7-8, a salt (NaCl, KCl or LiCl) in the range of 100-200 mM, a metal ion chelator (e.g., EDTA, diethylenetriaminepenta-acetic acid (DTPA), malate, inositol hexaphosphate, tripolyphosphate or polyphosphoric acid), a nonreducing free radical scavenger (e.g. ethanol, glycerol, methionine or dimethyl sulfoxide) and the highest appropriate DNA concentration in a sterile glass vial, packaged to protect the highly purified, nuclease free DNA from light. A particularly preferred formulation which will enhance long term stability of the DNA vector vaccines of the present invention would comprise a Tris-HCl buffer at a pH from about 8.0 to about 9.0; ethanol or glycerol at about 3% w/v; EDTA or DTPA in a concentration range up to about 5 mM; and NaCl at a concentration from about 50 mM to about 500 mM. The use of such stabilized DNA vector vaccines and various alternatives to this preferred formulation range is described in detail in PCT International Application No. PCT/US97/06655 and PCT International Publication No. WO 97/40839, both of which are hereby incorporated by reference.

The DNA vector vaccines of the present invention may also be formulated with an adjuvant or adjuvants which may increase immunogenicity of the DNA polynucleotide vaccines of the present invention. A number of these adjuvants are known in the art and are available for use in a DNA vaccine, including but not limited to particle bombardment using DNA-coated gold beads, co-administration of DNA vaccines with plasmid DNA expressing cytokines, chemokines, or costimulatory molecules, formulation of DNA with cationic lipids or with experimental adjuvants such as saponin, monophosphoryl lipid A or other compounds which increase immunogenicity of the DNA vaccine. Another adjuvant for use in the DNA vector vaccines of the present invention are one or more forms of an aluminum phosphate-based adjuvant wherein the aluminum phosphate-based adjuvant possesses a molar PO₄/Al ratio of approximately 0.9. An additional mineral-based adjuvant may be generated from one or more forms of a calcium phosphate. These mineral-based adjuvants are useful in increasing cellular and humoral responses to DNA vaccination. These mineral-based compounds for use as DNA vaccines adjuvants are disclosed in PCT International

Application No. PCT/US98/02414, PCT International Publication No. WO 98/35562, which is hereby incorporated by reference. Another preferred adjuvant is a non-ionic block copolymer which shows adjuvant activity with DNA vaccines. The basic structure comprises blocks of polyoxyethylene (POE) and 5 polyoxypropylene (POP) such as a POE-POP-POE block copolymer. Newman et al. (1998, Critical Reviews in Therapeutic Drug Carrier Systems 15(2): 89-142) review a class of non-ionic block copolymers which show adjuvant activity. The basic structure comprises blocks of polyoxyethylene (POE) and polyoxypropylene (POP) such as a POE-POP-POE block copolymer. Newman et al. id., disclose 10 that certain POE-POP-POE block copolymers may be useful as adjuvants to an influenza protein-based vaccine, namely higher molecular weight POE-POP-POE block copolymers containing a central POP block having a molecular weight of over about 9000 daltons to about 20,000 daltons and flanking POE blocks which comprise up to about 20% of the total molecular weight of the copolymer (see also 15 U.S. Reissue Patent No. 36,665, U.S. Patent No. 5,567,859, U.S. Patent No. 5,691,387, U.S. Patent No. 5,696,298 and U.S. Patent No. 5,990,241, all issued to Emanuele, et al., regarding these POE-POP-POE block copolymers). WO 96/04932 further discloses higher molecular weight POE/POP block copolymers which have surfactant characteristics and show biological efficacy as 20 vaccine adjuvants. The above cited references within this paragraph are hereby incorporated by reference in their entirety. It is therefore within the purview of the skilled artisan to utilize available adjuvants which may increase the immune response of the polynucleotide vaccines of the present invention in comparison to administration of a non-adjuvanted polynucleotide vaccine.

The DNA vector vaccines of the present invention are administered to the host by any means known in the art, such as enteral and parenteral routes. These routes of delivery include but are not limited to intramusclar injection, intraperitoneal injection, intravenous injection, inhalation or intranasal delivery, oral delivery, sublingual administration, subcutaneous administration, transdermal administration, transcutaneous administration, percutaneous administration or any form of particle bombardment, such as a biolostic device such as a "gene gun" or by any available needle-free injection device. The preferred methods of delivery of the HIV-1 Polbased DNA vaccines disclosed herein are intramuscular injection, subcutaneous administration and needle-free injection. An especially preferred method is

25

30

intramuscular delivery.

5

10

15

20

25

30

The amount of expressible DNA to be introduced to a vaccine recipient will depend on the strength of the transcriptional and translational promoters used in the DNA construct, and on the immunogenicity of the expressed gene product. In general, an immunologically or prophylactically effective dose of about 1 μ g to greater than about 20 mg, and preferably in doses from about 1 mg to about 5 mg is administered directly into muscle tissue. As noted above, subcutaneous injection, intradermal introduction, impression through the skin, and other modes of administration such as intraperitoneal, intravenous, inhalation and oral delivery are also contemplated. It is also contemplated that booster vaccinations are to be provided in a fashion which optimizes the overall immune response to the Pol-based DNA vector vaccines of the present invention.

The aforementioned polynucleotides, when directly introduced into a vertebrate *in vivo*, express the respective HIV-1 Pol protein within the animal and in turn induce a cellular immune response within the host to the expressed Pol antigen. To this end, the present invention also relates to methods of using the HIV-1 Polbased polynucleotide vaccines of the present invention to provide effective immunoprophylaxis, to prevent establishment of an HIV-1 infection following exposure to this virus, or as a post-HIV infection therapeutic vaccine to mitigate the acute HIV-1 infection so as to result in the establishment of a lower virus load with beneficial long term consequences. As noted above, the present invention contemplates a method of administration or use of the DNA pol-based vaccines of the present invention using an any of the known routes of introducing polynucleotides into living tissue to induce expression of proteins.

Therefore, the present invention provides for methods of using a DNA pol-based vaccine utilizing the various parameters disclosed herein as well as any additional parameters known in the art, which, upon introduction into mammalian tissue induces intracellular expression of these DNA pol-based vaccines. This intracellular expression of the Pol-based immunogen induces a cellular immune response which provides a substantial level of protection against an existing HIV-1 infection or provides a substantial level of protection against a future infection in a presently uninfected host.

The following examples are provided to illustrate the present invention without, however, limiting the same hereto.

EXAMPLE 1

Vaccine Vectors

VI - Vaccine vector V1 was constructed from pCMVIE-AKI-DHFR (Whang et al., 1987, J. Virol. 61: 1796). The AKI and DHFR genes were removed by cutting the vector with EcoRI and self-ligating. This vector does not contain intron A in the CMV promoter, so it was added as a PCR fragment that had a deleted internal SacI site [at 1855 as numbered in Chapman, et al., 1991, Nuc. Acids Res. 19: 3979). The template used for the PCR reactions was pCMVintA-Lux, made by ligating the 10 HindIII and NheI fragment from pCMV6a120 (see Chapman et al., ibid.), which includes hCMV-IE1 enhancer/promoter and intron A, into the HindIII and XbaI sites of pBL3 to generate pCMVIntBL. The 1881 base pair luciferase gene fragment (HindIII-SmaI Klenow filled-in) from RSV-Lux (de Wet et al., 1987, Mol. Cell Biol. 7: 725) was ligated into the Sall site of pCMVIntBL, which was Klenow filled-in and 15 phosphatase treated. The primers that spanned intron A are: 5' primer: 5'-CTATAT AAGCAGAGCTCGTTTAG-3' (SEQ ID NO:10); 3' primer: 5'-GTAGCAAA GATCTAAGGACGGTGACTGCAG-3' (SEQ ID NO:11). The primers used to remove the SacI site are: sense primer, 5'-GTATGTGTCTGAAAATGAGCG TGGAGATTGGGCTCGCAC-3' (SEQ ID NO:12) and the antisense primer, 20 5'-GTGCGAGCCCAATCTCCACGCTCATTTTCAGAC ACATAC-3' (SEQ ID NO:13). The PCR fragment was cut with Sac I and Bgl II and inserted into the vector which had been cut with the same enzymes.

VIJ – Vaccine vector V1J was generated to remove the promoter and transcription termination elements from vector V1 in order to place them within a more defined context, create a more compact vector, and to improve plasmid purification yields. V1J is derived from vectors V1 and pUC18, a commercially available plasmid. V1 was digested with SspI and EcoRI restriction enzymes producing two fragments of DNA. The smaller of these fragments, containing the CMVintA promoter and Bovine Growth Hormone (BGH) transcription termination elements which control the expression of heterologous genes, was purified from an agarose electrophoresis gel. The ends of this DNA fragment were then "blunted" using the T4 DNA polymerase enzyme in order to facilitate its ligation to another "blunt-ended" DNA fragment. pUC18 was chosen to provide the "backbone" of the expression vector. It is known to produce high yields of plasmid, is well-

25

PCT/US00/34724 WO 01/45748

characterized by sequence and function, and is of small size. The entire lac operon was removed from this vector by partial digestion with the HaeII restriction enzyme. The remaining plasmid was purified from an agarose electrophoresis gel, blunt-ended with the T4 DNA polymerase treated with calf intestinal alkaline phosphatase, and ligated to the CMVintA/BGH element described above. Plasmids exhibiting either of 5 two possible orientations of the promoter elements within the pUC backbone were obtained. One of these plasmids gave much higher yields of DNA in E. coli and was designated VIJ. This vector's structure was verified by sequence analysis of the junction regions and was subsequently demonstrated to give comparable or higher expression of heterologous genes compared with V1. The nucleotide sequence of V1J 10 is as follows: TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG CTATTGGCCA TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT ACATAACTTA CGGTAAATGG CCCGCCTGGC TGACCGCCCA ACGACCCCCG CCCATTGACG TCAATAATGA CGTATGTTCC CATAGTAACG CCAATAGGGA CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC 20 TGCCCACTTG GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCCTAC TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCGGT TTTGGCAGTA CATCAATGGG CGTGGATAGC GGTTTGACTC ACGGGGATTT CCAAGTCTCC ACCCCATTGA CGTCAATGGG AGTTTGTTTT GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA 25 CTCCGCCCCA TTGACGCAAA TGGGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG AGCTCGTTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACGCTGTT TTGACCTCCA TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA CGGTGCATTG GAACGCGGAT TCCCCGTGCC AAGAGTGACG TAAGTACCGC CTATAGAGTC TATAGGCCCCA CCCCCTTGGC TTCTTATGCA TGCTATACTG TTTTTGGCTT GGGGTCTATA CACCCCCGCT TCCTCATGTT 30 ATAGGTGATG GTATAGCTTA GCCTATAGGT GTGGGTTATT GACCATTATT GACCACTCCC CTATTGGTGA CGATACTTTC CATTACTAAT CCATAACATG GCTCTTTGCC ACAACTCTCT TTATTGGCTA TATGCCAATA CACTGTCCTT CAGAGACTGA CACGGACTCT GTATTTTAC AGGATGGGGT CTCATTTATT ATTTACAAAT TCACATATAC AACACCACCG TCCCCAGTGC

CCGCAGTTTT TATTAAACAT AACGTGGGAT CTCCACGCGA ATCTCGGGTA CGTGTTCCGG ACATGGGCTC TTCTCCGGTA GCGGCGGAGC TTCTACATCC GAGCCCTGCT CCCATGCCTC CAGCGACTCA TGGTCGCTCG GCAGCTCCTT GCTCCTAACA GTGGAGGCCA GACTTAGGCA CAGCACGATG CCCACCACCA CCAGTGTGCC GCACAAGGCC GTGGCGGTAG GGTATGTGTC TGAAAATGAG CTCGGGGAGC GGGCTTGCAC CGCTGACGCA TTTGGAAGAC TTAAGGCAGC GGCAGAAGAA GATGCAGGCA GCTGAGTTGT TGTGTTCTGA TAAGAGTCAG AGGTAACTCC CGTTGCGGTG CTGTTAACGG TGGAGGGCAG TGTAGTCTGA GCAGTACTCG TTGCTGCCGC GCGCGCCACC AGACATAATA GCTGACAGAC TAACAGACTG TTCCTTTCCA TGGGTCTTTT CTGCAGTCAC CGTCCTTAGA TCTGCTGTGC CTTCTAGTTG CCAGCCATCT GTTGTTTGCC CCTCCCCGT GCCTTCCTTG ACCCTGGAAG GTGCCACTCC CACTGTCCTT TCCTAATAAA 10 GGCAGCACAG CAAGGGGGAG GATTGGGAAG ACAATAGCAG GCATGCTGGG GATGCGGTGG GCTCTATGGG TACCCAGGTG CTGAAGAATT GACCCGGTTC CTCCTGGGCC AGAAAGAAGC AGGCACATCC CCTTCTCTGT GACACACCCT GTCCACGCCC CTGGTTCTTA GTTCCAGCCC 15 CACTCATAGG ACACTCATAG CTCAGGAGGG CTCCGCCTTC AATCCCACCC GCTAAAGTAC TTGGAGCGGT CTCTCCCTCC CTCATCAGCC CACCAAACCA AACCTAGCCT CCAAGAGTGG GAAGAAATTA AAGCAAGATA GGCTATTAAG TGCAGAGGGA GAGAAAATGC CTCCAACATG TGAGGAAGTA ATGAGAGAAA TCATAGAATT TCTTCCGCTT CCTCGCTCAC TGACTCGCTG CGCTCGGTCG TTCGGCTGCG GCGAGCGGTA TCAGCTCACT CAAAGGCGGT AATACGGTTA 20 TCCACAGAAT CAGGGGATAA CGCAGGAAAG AACATGTGAG CAAAAGGCCA GCAAAAGGCC AGGAACCGTA AAAAGGCCGC GTTGCTGGCG TTTTTCCATA GGCTCCGCCC CCCTGACGAG CATCACAAAA ATCGACGCTC AAGTCAGAGG TGGCGAAACC CGACAGGACT ATAAAGATAC CAGGCGTTTC CCCCTGGAAG CTCCCTCGTG CGCTCTCCTG TTCCGACCCT GCCGCTTACC GGATACCTGT CCGCCTTTCT CCCTTCGGGA AGCGTGGCGC TTTCTCAATG CTCACGCTGT 25 AGGTATCTCA GTTCGGTGTA GGTCGTTCGC TCCAAGCTGG GCTGTGTGCA CGAACCCCC GTTCAGCCCG ACCGCTGCGC CTTATCCGGT AACTATCGTC TTGAGTCCAA CCCGGTAAGA CACGACTTAT CGCCACTGGC AGCAGCCACT GGTAACAGGA TTAGCAGAGC GAGGTATGTA GGCGGTGCTA CAGAGTTCTT GAAGTGGTGG CCTAACTACG GCTACACTAG AAGGACAGTA TTTGGTATCT GCGCTCTGCT GAAGCCAGTT ACCTTCGGAA AAAGAGTTGG TAGCTCTTGA 30 TCCGGCAAAC AAACCACCGC TGGTAGCGGT GGTTTTTTTG TTTGCAAGCA GCAGATTACG CGCAGAAAAA AAGGATCTCA AGAAGATCCT TTGATCTTTT CTACGGGGTC TGACGCTCAG TGGAACGAAA ACTCACGTTA AGGGATTTTG GTCATGAGAT TATCAAAAAG GATCTTCACC TAGATCCTTT TAAATTAAAA ATGAAGTTTT AAATCAATCT AAAGTATATA TGAGTAAACT TGGTCTGACA GTTACCAATG CTTAATCAGT GAGGCACCTA TCTCAGCGAT CTGTCTATTT

CGTTCATCCA TAGTTGCCTG ACTCCCCGTC GTGTAGATAA CTACGATACG GGAGGGCTTA CCATCTGGCC CCAGTGCTGC AATGATACCG CGAGACCCAC GCTCACCGGC TCCAGATTTA TCAGCAATAA ACCAGCCAGC CGGAAGGGCC GAGCGCAGAA GTGGTCCTGC AACTTTATCC GCCTCCATCC AGTCTATTAA TTGTTGCCGG GAAGCTAGAG TAAGTAGTTC GCCAGTTAAT AGTTTGCGCA ACGTTGTTGC CATTGCTACA GGCATCGTGG TGTCACGCTC GTCGTTTGGT 5 ATGGCTTCAT TCAGCTCCGG TTCCCAACGA TCAAGGCGAG TTACATGATC CCCCATGTTG TGCAAAAAG CGGTTAGCTC CTTCGGTCCT CCGATCGTTG TCAGAAGTAA GTTGGCCGCA GTGTTATCAC TCATGGTTAT GGCAGCACTG CATAATTCTC TTACTGTCAT GCCATCCGTA AGATGCTTTT CTGTGACTGG TGAGTACTCA ACCAAGTCAT TCTGAGAATA GTGTATGCGG CGACCGAGTT GCTCTTGCCC GGCGTCAATA CGGGATAATA CCGCGCCACA TAGCAGAACT 10 TTAAAAGTGC TCATCATTGG AAAACGTTCT TCGGGGCGAA AACTCTCAAG GATCTTACCG CTGTTGAGAT CCAGTTCGAT GTAACCCACT CGTGCACCCA ACTGATCTTC AGCATCTTTT ACTTTCACCA GCGTTTCTGG GTGAGCAAAA ACAGGAAGGC AAAAATGCCGC AAAAAAGGGA ATAAGGGCGA CACGGAAATG TTGAATACTC ATACTCTTCC TTTTTCAATA TTATTGAAGC ATTTATCAGG GTTATTGTCT CATGAGCGGA TACATATTTG AATGTATTTA GAAAAATAAA 15 CAAATAGGGG TTCCGCGCAC ATTTCCCCGA AAAGTGCCAC CTGACGTCTA AGAAACCATT ATTATCATGA CATTAACCTA TAAAAATAGG CGTATCACGA GGCCCTTTCG TC (SEQ ID NO:14).

20

25

30

V1Ineo - Construction of vaccine vector V1Ineo expression vector involved removal of the ampr gene and insertion of the kanr gene (neomycin phosphotransferase). The ampr gene from the pUC backbone of V1J was removed by digestion with SspI and Eam1105I restriction enzymes. The remaining plasmid was purified by agarose gel electrophoresis, blunt-ended with T4 DNA polymerase, and then treated with calf intestinal alkaline phosphatase. The commercially available kanr gene, derived from transposon 903 and contained within the pUC4K plasmid, was excised using the PstI restriction enzyme, purified by agarose gel electrophoresis, and blunt-ended with T4 DNA polymerase. This fragment was ligated with the V1J backbone and plasmids with the kanr gene in either orientation were derived which were designated as V1Jneo #s 1 and 3. Each of these plasmids was confirmed by restriction enzyme digestion analysis, DNA sequencing of the junction regions, and was shown to produce similar quantities of plasmid as VIJ. Expression of heterologous gene products was also comparable to V1J for these V1Jneo vectors. V1Ineo#3, referred to as V1Ineo hereafter, was selected which contains the kanr gene in the same orientation as the ampr gene in V1J as the expression construct and

provides resistance to neomycin, kanamycin and G418. The nucleotide sequence of V1Jneo is as follows:

TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGGGGG TCAGCGGGTG 5 TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG CTATTGGCCA TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT ACATAACTTA CGGTAAATGG 10 CCCGCCTGGC TGACCGCCCA ACGACCCCCG CCCATTGACG TCAATAATGA CGTATGTTCC CATAGTAACG CCAATAGGGA CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC TGCCCACTTG GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCCTAC TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCGGT TTTGGCAGTA 15 CATCAATGGG CGTGGATAGC GGTTTGACTC ACGGGGATTT CCAAGTCTCC ACCCCATTGA CGTCAATGGG AGTTTGTTTT GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA CTCCGCCCA TTGACGCAAA TGGGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG AGCTCGTTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACGCTGTT TTGACCTCCA TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA CGGTGCATTG GAACGCGGAT 20 TCCCCGTGCC AAGAGTGACG TAAGTACCGC CTATAGAGTC TATAGGCCCA CCCCCTTGGC TTCTTATGCA TGCTATACTG TTTTTGGCTT GGGGTCTATA CACCCCGGCT TCCTCATGTT ATAGGTGATG GTATAGCTTA GCCTATAGGT GTGGGTTATT GACCATTATT GACCACTCCC CTATTGGTGA CGATACTTTC CATTACTAAT CCATAACATG GCTCTTTGCC ACAACTCTCT TTATTGGCTA TATGCCAATA CACTGTCCTT CAGAGACTGA CACGGACTCT GTATTTTTAC 25 AGGATGGGGT CTCATTTATT ATTTACAAAT TCACATATAC AACACCACCG TCCCCAGTGC CCGCAGTTTT TATTAAACAT AACGTGGGAT CTCCACGCGA ATCTCGGGTA CGTGTTCCGG ACATGGGCTC TTCTCCGGTA GCGGCGGAGC TTCTACATCC GAGCCCTGCT CCCATGCCTC CAGCGACTCA TGGTCGCTCG GCAGCTCCTT GCTCCTAACA GTGGAGGCCA GACTTAGGCA CAGCACGATG CCCACCACCA CCAGTGTGCC GCACAAGGCC GTGGCGGTAG GGTATGTGTC 30 TGAAAATGAG CTCGGGGAGC GGGCTTGCAC CGCTGACGCA TTTGGAAGAC TTAAGGCAGC GGCAGAAGAA GATGCAGGCA GCTGAGTTGT TGTGTTCTGA TAAGAGTCAG AGGTAACTCC CGTTGCGGTG CTGTTAACGG TGGAGGGCAG TGTAGTCTGA GCAGTACTCG TTGCTGCCGC GCGCGCCACC AGACATAATA GCTGACAGAC TAACAGACTG TTCCTTTCCA TGGGTCTTTT CTGCAGTCAC CGTCCTTAGA TCTGCTGTGC CTTCTAGTTG CCAGCCATCT GTTGTTTGCC

	CCTCCCCCGT G	CCTTCCTTG A	ACCCTGGAAG	GTGCCACTCC	CACTGTCCTT	TCCTAATAAA
	ATGAGGAAAT T	GCATCGCAT '	TGTCTGAGTA	GGTGTCATTC	TATTCTGGGG	GGTGGGGTGG
	GGCAGCACAG C	AAGGGGGAG	GATTGGGAAG	ACAATAGCAG	GCATGCTGGG	GATGCGGTGG
	GCTCTATGGG T	ACCCAGGTG	CTGAAGAATT	GACCCGGTTC	CTCCTGGGCC	AGAAAGAAGC
5	AGGCACATCC C	сттстстст	GACACACCCT	GTCCACGCCC	CTGGTTCTTA	GTTCCAGCCC
5	CACTCATAGG A	CACTCATAG	CTCAGGAGGG	CTCCGCCTTC	AATCCCACCC	GCTAAAGTAC
	TTGGAGCGGT C	TCTCCCTCC	CTCATCAGCC	CACCAAACCA	AACCTAGCCT	CCAAGAGTGG
	GAAGAAATTA A	AGCAAGATA	GGCTATTAAG	TGCAGAGGGA	GAGAAAATGC	CTCCAACATG
	TGAGGAAGTA A	TGAGAGAAA	TCATAGAATT	TCTTCCGCTT	CCTCGCTCAC	TGACTCGCTG
10	CGCTCGGTCG	TCGGCTGCG	GCGAGCGGTA	TCAGCTCACT	CAAAGGCGGT	AATACGGTTA
10	TCCACAGAAT (CAGGGGATAA	CGCAGGAAAG	AACATGTGAG	CAAAAGGCCA	GCAAAAGGCC
•	AGGAACCGTA	AAAAGGCCGC	GTTGCTGGCG	TTTTTCCATA	GGCTCCGCCC	CCCTGACGAG
	САТСАСАААА	ATCGACGCTC	AAGTCAGAGG	TGGCGAAACC	CGACAGGACT	ATAAAGATAC
	CAGGCGTTTC	CCCCTGGAAG	CTCCCTCGTG	CGCTCTCCTG	TTCCGACCCI	GCCGCTTACC
15	GGATACCTGT	CCGCCTTTCT	CCCTTCGGGA	AGCGTGGCGC	TTTCTCAATG	CTCACGCTGT
	AGGTATCTCA	GTTCGGTGTA	GGTCGTTCGC	TCCAAGCTGG	GCTGTGTGCA	CGAACCCCCC
	GTTCAGCCCG	ACCGCTGCGC	CTTATCCGGT	AACTATCGTC	TTGAGTCCA	A CCCGGTAAGA
	CACGACTTAT	CGCCACTGGC	AGCAGCCACT	GGTAACAGG	TTAGCAGAGC	GAGGTATGTA
	GCCGGTGCTA	CAGAGTTCTT	GAAGTGGTG	CCTAACTACC	G GCTACACTAC	B AAGGACAGTA
20	TTTGGTATCT	GCGCTCTGCT	GAAGCCAGT'	r ACCTTCGGA	A AAAGAGTTG	G TAGCTCTTGA
	TCCGGCAAAC	AAACCACCGC	TGGTAGCGG'	r ggttttttt	G TTTGCAAGC	A GCAGATTACG
	CGCAGAAAAA	AAGGATCTCA	AGAAGATCC'	r TTGATCTTT	T CTACGGGGT	C TGACGCTCAG
	TGGAACGAAA	ACTCACGTTA	AGGGATTTT	G GTCATGAGA	TATCAAAAA	G GATCTTCACC
	TAGATCCTTT	TAAATTAAAT	ATGAAGTTT	T AAATCAATC	T AAAGTATAT	A TGAGTAAACT
25	TGGTCTGACA	GTTACCAATO	CTTAATCAG	T GAGGCACCT	A TCTCAGCGA	T CTGTCTATTT
	CGTTCATCCA	TAGTTGCCTC	ACTCCGGGG	g gggggggg	C TGAGGTCTG	C CTCGTGAAGA
	AGGTGTTGCT	GACTCATAC	C AGGCCTGAA	T CGCCCCATC	A TCCAGCCAG	A AAGTGAGGGA
	GCCACGGTTG	ATGAGAGCT	TGTTGTAGO	T GGACCAGTI	G GTGATTTTG	A ACTTTTGCTT
	TGCCACGGAA	CGGTCTGCG	r TGTCGGGAA	G ATGCGTGAT	C TGATCCTTC	A ACTCAGCAAA
30	AGTTCGATTT	ATTCAACAA	A GCCGCCGTC	C CGTCAAGTC	A GCGTAATGC	CT CTGCCAGTGT
	TACAACCAAT	TAACCAATT	C TGATTAGA	A AACTCATCO	SA GCATCAAA1	G AAACTGCAAT
	TTATTCATAT	CAGGATTAT	C AATACCATA	Y TTTTGAAA!	AA GCCGTTTCT	G TAATGAAGGA
	GAAAACTCAC	CGAGGCAGT	T CCATAGGA	rg gcaagatco	CT GGTATCGG	TGCGATTCCG
	ACTCGTCCAA	CATCAATAC	A ACCTATTA	AT TTCCCCTC	GT CAAAAATA	AG GTTATCAAGT

GAGAAATCAC CATGAGTGAC GACTGAATCC GGTGAGAATG GCAAAAGCTT ATGCATTTCT TTCCAGACTT GTTCAACAGG CCAGCCATTA CGCTCGTCAT CAAAATCACT CGCATCAACC AAACCGTTAT TCATTCGTGA TTGCGCCTGA GCGAGACGAA ATACGCGATC GCTGTTAAAA GGACAATTAC AAACAGGAAT CGAATGCAAC CGGCGCAGGA ACACTGCCAG CGCATCAACA 5 ATATTTTCAC CTGAATCAGG ATATTCTTCT AATACCTGGA ATGCTGTTTT CCCGGGGATC GCAGTGGTGA GTAACCATGC ATCATCAGGA GTACGGATAA AATGCTTGAT GGTCGGAAGA GGCATAAATT CCGTCAGCCA GTTTAGTCTG ACCATCTCAT CTGTAACATC ATTGGCAACG CTACCTTTGC CATGTTTCAG AAACAACTCT GGCGCATCGG GCTTCCCATA CAATCGATAG ATTGTCGCAC CTGATTGCCC GACATTATCG CGAGCCCATT TATACCCATA TAAATCAGCA 10 TCCATGTTGG AATTTAATCG CGGCCTCGAG CAAGACGTTT CCCGTTGAAT ATGGCTCATA ACACCCCTTG TATTACTGTT TATGTAAGCA GACAGTTTTA TTGTTCATGA TGATATATTT TTATCTTGTG CAATGTAACA TCAGAGATTT TGAGACACAA CGTGGCTTTC CCCCCCCCC CATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC 15 TAAGAAACCA TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT CGTC (SEQ ID NO:15).

VIJns - The expression vector VIJns was generated by adding an SfiI site to V1Jneo to facilitate integration studies. A commercially available 13 base pair SfiI linker (New England BioLabs) was added at the KpnI site within the BGH sequence of the vector. V1Jneo was linearized with KpnI, gel purified, blunted by T4 DNA polymerase, and ligated to the blunt SfiI linker. Clonal isolates were chosen by restriction mapping and verified by sequencing through the linker. The new vector was designated V1Jns. Expression of heterologous genes in V1Jns (with SfiI) was comparable to expression of the same genes in V1Jneo (with KpnI).

The nucleotide sequence of V1Jns is as follows:

20

25

30

TCGCCCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCG GAGACGGTCA
CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGGCGC TCAGCGGGTG
TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG
CTATTGGCCA TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG
TCCAACATTA CCGCCATGTT GACATTGAT ATTGACTAGT TATTAATAGT AATCAATTAC
GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT ACATAACTTA CGGTAAATGG
CCCGCCTGGC TGACCGCCCA ACGACCCCCG CCCATTGACG TCAATAATGA CGTATGTTCC
CATAGTAACG CCAATAGGGA CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC

PCT/US00/34724

	TGCCCACTTG GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
	TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCCTAC
	TTGGCAGTAA TGGCCCGCCT GGCATTATOC TATTACCATG GTGATGCGGT TTTGGCAGTA
	CATCAATGGG CGTGGATAGC GGTTTGACTC ACGGGGATTT CCAAGTCTCC ACCCCATTGA
_	CATCAATGGG CGTGGATAGC GGTTTGATGT TO CAACGGGAC TTTCCAAAAT GTCGTAACAA
5	CTCCGCCCCA TTGACGCAAA TGGGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG
	AGCTCGTTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACGCTGTT TTGACCTCCA
	AGCTCGTTTA GTGAACCGTC AGATCGCGTC TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA CGGTGCATTG GAACGCGGAT
	TAGAAGACAC CGGGACCGAT CCAGGGTGGG TATAGACTC TATAGGCACA CCCCTTTGGC TCCCCGTGCC AAGAGTGACG TAAGTACCGC CTATAGACTC TATAGGCACA CCCCTTTGGC
	TCCCCGTGCC AAGAGTGACG TAAGTACCCC COTTATGCTA TCTTATGCAT GCTATACTGT TTTTGGCTTG GGGCCTATAC ACCCCCGCTT CCTTATGCTA
10	TCTTATGCAT GCTATACTGT TITTGGCTTO COCCUTATATTG ACCACTATTG ACCACTCCCC TAGGTGATGG TATAGCTTAG CCTATAGGTG TGGGTTATTG ACCACTATTG ACCACTCCCC
	TAGGTGATGG TATAGCTTAG CCTATAGGTG TOOCHTAGGTG TATTGGTGAC GATACTTTCC ATTACTAATC CATAACATGG CTCTTTGCCA CAACTATCTC
	TATTGGTGAC GATACTITCC ATTACTANTO CONTROL ACGGACTCTG TATTTTTACA TATTGGCTAT ATGCCAATAC TCTGTCCTTC AGAGACTGAC ACGGACTCTG TATTTTTACA
	GGATGGGGTC CCATTTATTA TTTACAAATT CACATATACA ACAACGCCGT CCCCCGTGCC
_	GGATGGGGTC CCATTTATTA TITACAAATT GAGATGGGAA TCTCGGGTAC GTGTTCCGGA CGCAGTTTTT ATTAAACATA GCGTGGGATC TCCACGCGAA TCTCGGGTAC GTGTTCCGGA
15	CGCAGTTTTT ATTAAACATA GCGIGGGAIC TOCACATCCG AGCCCTGGTC CCATGCCTCC CATGGGCTCT TCTCCGGTAG CGGCGGAGCT TCCACATCCG AGCCCTGGTC CCATGCCTCC
	CATGGGCTCT TCTCCGGTAG CGGCGGAGGT TOSITOSTON AGCGGCTCAT GGTCGCTCGG CAGCTCCTTG CTCCTAACAG TGGAGGCCAG ACTTAGGCAC
	AGCGGCTCAT GGTCGCTCGG CAGCTCCTTG CTCCTTGTCTCT AGCACAATGC CCACCACCAC CAGTGTGCCG CACAAGGCCG TGGCGGTAGG GTATGTGTCT
	AGCACAATGC CCACCAC CAGIGIGGCG CITCLES CONTROL TAAGGCAGCG GAAAATGAGC GTGGAGATTG GGCTCGCACG GCTGACGCAG ATGGAAGACT TAAGGCAGCG
	GAAAATGAGC GTGGAGATTG GGCTCGCACG GCTCACGCTCGCACGCGCGCGCGCGCGCGCGCGCGCGCGC
20	GCAGAAGAAG ATGCAGGCAG CTGAGTTGTT GTMTTGTGTGTGTGTGTGCGGCGGGGGTTGCTGCGGGGGGGG
	GTTGCGGTGC TGTTAACGGT GGAGGGCAGT GTTATCCAT GGGTCTTTTC CGCGCCACCA GACATAATAG CTGACAGACT AACAGACTGT TCCTTTCCAT GGGTCTTTTC
	CGCGCCACCA GACATAATAG CTGACAGACT AACAGATGT CAGCCATCTG TTGTTTGCCC TGCAGTCACC GTCCTTAGAT CTGCTGTGCC TTCTAGTTGC CAGCCATCTG TTGTTTGCCC
	TGCAGTCACC GTCCTTAGAT CTGCTGTGCC TTCTMGTTGC CTCCCCCGTG CCTTCCTTGA CCCTGGAAGG TGCCACTCCC ACTGTCCTTT CCTAATAAAA
	THE STATE OF THE PROPERTY OF T
25 .	TGAGGAAATT GCATCGCATT GTCTGAGTAG GTGTCATTO2 TATAGCAGG ATGCGGTGGG GCAGGACAGC AAGGGGGAGG ATTGGGAAGA CAATAGCAGG CATGCTGGGG ATGCGGTGGG
	GCAGGACAGC AAGGGGGAGG ATTGGGAAGA CAATMOONO CONTROL CON
	CTCTATGGCC GCTGCGGCCA GGTGCTGAAG AATTGACGG GCCCCTGGTT CTTAGTTCCA AAGCAGGCAC ATCCCCTTCT CTGTGACACA CCCTGTCCAC GCCCCTGGTT CTTAGTTCCA
	AAGCAGGCAC ATCCCCTTCT CTGTGACACA CCCTGTGGTG GCCCGCTAAA GCCCCACTCA TAGGACACTC ATAGCTCAGG AGGGCTCCGC CTTCAATCCC ACCCGCTAAA
	GCCCCACTCA TAGGACACTC ATAGCTCAGG AGGGCTCCCOA ACCAAACCTA GCCTCCAAGA
30	GTACTTGGAG CGGTCTCTCC CTCCCTCATC AGCCCACCAA ACCAAACCTA GCCTCCAAGA GTGGGAAGAA ATTAAAGCAA GATAGGCTAT TAAGTGCAGA GGGAGAGAAA ATGCCTCCAA
•	GTGGGAAGAA ATTAAAGCAA GATAGGCTAT TAAGTGCAGA COCTCCTCGC TCACTGACTC
	CATGTGAGGA AGTAATGAGA GAAATCATAG AATTTCTTCC GCTTCCTCGC TCACTGACTC
	GCTGCGCTCG GTCGTTCGGC TGCGGCGAGC GGTATCAGCT CACTCAAAGG CGGTAATACG
	GTTATCCACA GAATCAGGGG ATAACGCAGG AAAGAACATG TGAGCAAAAAG GCCAGCAAAA

	GGCCAGGAAC	CGTAAAAAGG	CCGCGTTGCT	GGCGTTTTTC	CATAGGCTCC	GCCCCCTGA
	CGAGCATCAC	AAAAATCGAC	GCTCAAGTCA	GAGGTGGCGA	AACCCGACAG	GACTATAAAG
	ATACCAGGCG	TTTCCCCCTG	GAAGCTCCCT	CGTGCGCTCT	CCTGTTCCGA	CCCTGCCGCT
	TACCGGATAC	CTGTCCGCCT	TTCTCCCTTC	GGGAAGCGTG	GCGCTTTCTC	ATAGCTCACG
5	CTGTAGGTAT	CTCAGTTCGG	TGTAGGTCGT	TCGCTCCAAG	CTGGGCTGTG	TGCACGAACC
	CCCCGTTCAG	CCCGACCGCT	GCGCCTTATC	CGGTAACTAT	CGTCTTGAGT	CCAACCCGGT
	AAGACACGAC	TTATCGCCAC	TGGCAGCAGC	CACTGGTAAC	AGGATTAGCA	GAGCGAGGTA
	TGTAGGCGGT	GCTACAGAGT	TCTTGAAGTG	GTGGCCTÄAC	TACGGCTACA	CTAGAAGAAC
	AGTATTTGGT	ATCTGCGCTC	TGCTGAAGCC	AGTTACCTTC	GGAAAAAGAG	TTGGTAGCTC
10	TTGATCCGGC	AAACAAACCA	CCGCTGGTAG	CGGTGGTTTT	TTTGTTTGCA	AGCAGCAGAT
	TACGCGCAGA	AAAAAAGGAT	CTCAAGAAGA	TCCTTTGATC	TTTTCTACGG	GGTCTGACGC
	TCAGTGGAAC	GAAAACTCAC	GTTAAGGGAT	TTTGGTCATG	AGA [†] TATCAA	AAAGGATCTT
	CACCTAGATC	CTTTTAAATT	AAAAATGAAG	TTTTAAATCA	ATCTAAAGTA	TATATGAGTA
	AACTTGGTCT	GACAGTTACC	AATGCTTAAT	CAGTGAGGCA	CCTATCTCAG	CGATCTGTCT
15	ATTTCGTTCA	TCCATAGTTG	CCTGACTCGG	GGGGGGGGG	CGCTGAGGTC	TGCCTCGTGA
	AGAAGGTGTT	GCTGACTCAT	ACCAGGCCTG	AATCGCCCCA	TCATCCAGCC	AGAAAGTGAG
	GGAGCCACGG	TTGATGAGAG	CTTTGTTGTA	GGTGGACCAG	TTGGTGATTT	TGAACTTTTG
	CTTTGCCACG	GAACGGTCTG	CGTTGTCGGG	AAGATGCGTG	ATCTGATCCT	TCAACTCAGC
	AAAAGTTCGA	TTTATTCAAC	AAAGCCGCCG	TCCCGTCAAG	TCAGCGTAAT	GCTCTGCCAG
20	TGTTACAACC	AATTAACCAA	TTCTGATTAG	AAAAACTCAT	CGAGCATCAA	ATGAAACTGC
	AATTTATTCA	TATCAGGATT	ATCAATACCA	TATTTTTGAA	AAAGCCGTTT	CTGTAATGAA
	GGAGAAAACT	CACCGAGGCA	GTTCCATAGG	ATGGCAAGAT	CCTGGTATCG	GTCTGCGATT
	CCGACTCGTC	CAACATCAAT	ACAACCTATT	AATTTCCCCT	CGTCAAAAAT	AAGGTTATCA
	AGTGAGAAAT	CACCATGAGT	GACGACTGAA	TCCGGTGAGA	ATGGCAAAAG	CTTATGCATT
25	TCTTTCCAGA	CTTGTTCAAC	AGGCCAGCCA	TTACGCTCGT	CATCAAAATC	ACTCGCATCA
	ACCAAACCGT	TATTCATTCG	TGATTGCGCC	TGAGCGAGAC	GAAATACGCG	ATCGCTGTTA
	AAAGGACAAT	TACAAACAGG	AATCGAATGC	AACCGGCGCA	GGAACACTGC	CAGCGCATCA
	ACAATATTTT	CACCTGAATC	AGGATATTCT	TCTAATACCT	GGAATGCTGT	TTTCCCGGGG
	ATCGCAGTGG	TGAGTAACCA	TGCATCATCA	GGAGTACGGA	TAAAATGCTT	GATGGTCGGA
30	AGAGGCATAA	ATTCCGTCAG	CCAGTTTAGT	CTGACCATCT	CATCTGTAAC	ATCATTGGCA
	ACGCTACCTT	TGCCATGTTT	CAGAAACAAC	TCTGGCGCAT	CGGGCTTCCC	ATACAATCGA
	TAGATTGTCG	CACCTGATTG	CCCGACATTA	TCGCGAGCCC	ATTTATACCC	ATATAAATCA
	GCATCCATGT	TGGAATTTAA	TCGCGGCCTC	GAGCAAGACG	TTTCCCGTTG	AATATGGCTC
	ATAACACCCC	TTGTATTACT	GTTTATGTAA	GCAGACAGTT	TTATTGTTCA	TGATGATATA

TTTTTATCTT GTGCAATGTA ACATCAGAGA TTTTGAGACA CAACGTGGCT TTCCCCCCC CCCCATTATT GAAGCAATTA TCAGGGTTAT TGTCTCATGA GCGGATACAT ATTTGAATGT ATTTAGAAAA ATAAACAAAT AGGGGTTCCG CGCACATTTC CCCGAAAAGT GCCACCTGAC GTCTAAGAAA CCATTATTAT CATGACATTA ACCTATAAAA ATAGGCGTAT CACGAGGCCC TTTCCGTC(SEQ ID NO:16).

The underlined nucleotides of SEQ ID NO:16 represent the Sfi1 site introduced into the Kpn 1 site of V1Ineo.

5

VIJns-tPA - The vaccine vector V1Jns-tPA was constructed in order to fuse an heterologous leader peptide sequence to the pol DNA constructs of the present invention. More specifically, the vaccine vector V1Jns was modified to include the 10 human tissue-specific plasminogen activator (tPA) leader. As an exemplification, but by no means a limitation of generating a pol DNA construct comprising an aminoterminal leader sequence, plasmid V1Ineo was modified to include the human tissuespecific plasminogen activator (tPA) leader. Two synthetic complementary oligomers were annealed and then ligated into V1Ineo which had been BglII digested. The 15 sense and antisense oligomers were 5'-GATCACCATGGATGCAATGAAGAG AGGGCTCTGCTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCAG CGA-3' (SEQ ID NO:17); and, 5'-GATCTCGCTGGGCGAAACGAAGACTGCTCC ACACAGCAGCACACAGCAGAGCCCTCTCTTCATTGCATCCATGGT-3' (SEQ ID NO:18). The Kozak sequence is underlined in the sense oligomer. These 20 oligomers have overhanging bases compatible for ligation to BgIII-cleaved sequences. After ligation the upstream BglII site is destroyed while the downstream BglII is retained for subsequent ligations. Both the junction sites as well as the entire tPA leader sequence were verified by DNA sequencing. Additionally, in order to conform with V1Ins (=V1Ineo with an SfiI site), an SfiI restriction site was placed at the KpnI 25 site within the BGH terminator region of V1Ineo-tPA by blunting the KpnI site with T4 DNA polymerase followed by ligation with an SfiI linker (catalogue #1138, New England Biolabs), resulting in V1Jns-tPA. This modification was verified by restriction digestion and agarose gel electrophoresis.

The V1Ins-tpa vector nucleotide sequence is as follows:

TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA
CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG
TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG

	CTATTGGCCA	TTGCATACGT	TGTATCCATA	TCATAATATG	TACATTTATA	TTGGCTCATG
_	TCCAACATTA	CCGCCATGTT	GACATTGATT	ATTGACTAGT	TATTAATAGT	AATCAATTAC
	GGGGTCATTA	GTTCATAGCC	CATATATGGA	GTTCCGCGTT	ACATAACTTA	CGGTAAATGG
	CCCGCCTGGC	TGACCGCCCA	ACGACCCCCG	CCCATTGACG	TCAATAATGA	CGTATGTTCC
5	CATAGTAACG	CCAATAGGGA	CTTTCCATTG	ACGTCAATGG	GTGGAGTATT	TACGGTAAAC
	TGCCCACTTG	GCAGTACATC	AAGTGTATCA	TATGCCAAGT	ACGCCCCTA	TTGACGTCAA
	TGACGGTAAA	TGGCCCGCCT	GGCATTATGC	CCAGTACATG	ACCTTATGGG	ACTTTCCTAC
	TTGGCAGTAC	ATCTACGTAT	TAGTCATCGC	TATTACCATG	GTGATGCGGT	TTTGGCAGTA
	CATCAATGGG	CGTGGATAGC	GGTTTGACTC	ACGGGGATTT	CCAAGTCTCC	ACCCCATTGA
10	CGTCAATGGG	AGTTTGTTTT	GGCACCAAAA	TCAACGGGAC	TTTCCAAAAT	GTCGTAACAA
	CTCCGCCCCA	TTGACGCAAA	TGGGCGGTAG	GCGTGTACGG	TGGGAGGTCT	ATATAAGCAG
	AGCTCGTTTA	GTGAACCGTC	AGATCGCCTG	GAGACGCCAT	CCACGCTGTT	TTGACCTCCA
	TAGAAGACAC	CGGGACCGAT	CCAGCCTCCG	CGGCCGGGAA	CGGTGCATTG	GAACGCGGAT
	TCCCCGTGCC	AAGAGTGACG	TAAGTACCGC	CTATAGACTC	TATAGGCACA	CCCCTTTGGC
15	TCTTATGCAT	GCTATACTGT	TTTTGGCTTG	GGGCCTATAC	ACCCCCGCTT	CCTTATGCTA
	TAGGTGATGG	TATAGCTTAG	CCTATAGGTG	TGGGTTATTG	ACCATTATTG	ACCACTCCCC
	TATTGGTGAC	GATACTTTCC	ATTACTAATC	CATAACATGG	CTCTTTGCCA	CAACTATCTC
	TATTGGCTAT	ATGCCAATAC	TCTGTCCTTC	AGAGACTGAC	ACGGACTCTG	TATTTTTACA
	GGATGGGGTC	CCATTTATTA	TTTACAAATT	CACATATACA	ACAACGCCGT	CCCCCGTGCC
20	CGCAGTTTTT	ATTAAACATA	GCGTGGGATC	TCCACGCGAA	TCTCGGGTAC	GTGTTCCGGA
	CATGGGCTCT	TCTCCGGTAG	CGGCGGAGCT	TCCACATCCG	AGCCCTGGTC	CCATGCCTCC
	AGCGGCTCAT	GGTCGCTCGG	CAGCTCCTTG	CTCCTAACAG	TGGAGGCCAG	ACTTAGGCAC
	AGCACAATGC	CCACCACCAC	CAGTGTGCCG	CACAAGGCCG	TGGCGGTAGG	GTATGTGTCT
	GAAAATGAGC	GTGGAGATTG	GGCTCGCACG	GCTGACGCAG	ATGGAAGACT	TAAGGCAGCG
25	GCAGAAGAAG	ATGCAGGCAG.	CTGAGTTGTT	GTATTCTGAT	AAGAGTCAGA	GGTAACTCCC ·
	GTTGCGGTGC	TGTTAACGGT	GGAGGGCAGT	GTAGTCTGAG	CAGTACTCGT	TGCTGCCGCG
	CGCGCCACCA	GACATAATAG	CTGACAGACT	AACAGACTGT	TCCTTTCCAT	GGGTCTTTTC
		GTCCTT <u>AGAT</u>				
	-	GAGCAGTCTT				
30	GCCATCTGTT	GTTTGCCCCT	CCCCCGTGCC	TTCCTTGACC	CTGGAAGGTG	CCACTCCCAC
	TGTCCTTTCC	TAATAAAATG	AGGAAATTGC	ATCGCATTGT	CTGAGTAGGT	GTCATTCTAT
		GGGGTGGGGC				
		GCGGTGGGCT				
	TCCTCCTGGG	CCAGAAAGAA	GCAGGCACAT	CCCCTTCTCT	GTGACACACC	CTGTCCACGC

	CCCTGGTTCT	TAGTTCCAGC	CCCACTCATA	GGACACTCAT	AGCTCAGGAG	GGCTCCGCCT
	TCAATCCCAC	CCGCTAAAGT	ACTTGGAGCG	GTCTCTCCCT	CCCTCATCAG	CCCACCAAAC
	CAAACCTAGC	CTCCAAGAGT	GGGAAGAAAT	TAAAGCAAGA	TAGGCTATTA	AGTGCAGAGG
	CACAGAAAAT	GCCTCCAACA	TGTGAGGAAG	TAATGAGAGA	AATCATAGAA	TTTCTTCCGC
5	TTCCTCGCTC	ACTGACTCGC	TGCGCTCGGT	CGTTCGGCTG	CGGCGAGCGG	TATCAGCTCA
,	CTCAAAGGCG	GTAATACGGT	TATCCACAGA	ATCAGGGGAT	AACGCAGGAA	AGAACATGTG
	AGCAAAAGGC	CAGCAAAAGG	CCAGGAACCG	TAAAAAGGCC	GCGTTGCTGG	CGTTTTTCCA
	TAGGCTCCGC	CCCCCTGACG	AGCATCACAA	AAATCGACGC	TCAAGTCAGA	GGTGGCGAAA
	CCCGACAGGA	CTATAAAGAT	ACCAGGCGTT	TCCCCCTGGA	AGCTCCCTCG	TGCGCTCTCC
10	TGTTCCGACC	CTGCCGCTTA	CCGGATACCT	GTCCGCCTTT	CTCCCTTCGG	GAAGCGTGGC
•	GCTTTCTCAT	AGCTCACGCT	GTAGGTATCT	CAGTTCGGTG	TAGGTCGTTC	GCTCCAAGCT
	GGGCTGTGTG	CACGAACCCC	CCGTTCAGCC	CGACCGCTGC	GCCTTATCCG	GTAACTATCG
	TCTTGAGTCC	AACCCGGTAA	GACACGACTT	ATCGCCACTG	GCAGCAGCCA	CTGGTAACAG
	GATTAGCAG	A GCGAGGTATG	TAGGCGGTGC	TACAGAGTTC	TTGAAGTGGT	GGCCTAACTA
15	CGGCTACACT	r agaagaacag	TATTTGGTAT	CTGCGCTCTG	CTGAAGCCAG	TTACCTTCGG
	AAAAAGAGT'	r ggtagctctt	GATCCGGCA	ACAAACCACC	GCTGGTAGCG	GTGGTTTTTT
	TGTTTGCAA	G CAGCAGATTA	CGCGCAGAA	AAAAGGATCT	CAAGAAGATC	CTTTGATCTT
	TTCTACGGG	G TCTGACGCTC	AGTGGAACG	AAACTCACGT	TAAGGGATTT	TGGTCATGAG
	ATTATCAAA	A AGGATCTTCA	CCTAGATCC	AATTAAATTA	AAATGAAGTT	TTAAATCAAT
20	CTAAAGTAT	A TATGAGTAAA	CTTGGTCTG	A CAGTTACCAA	TGCTTAATCA	GTGAGGCACC
	TATCTCAGC	G ATCTGTCTAT	TTCGTTCAT	CATAGTTGCC	TGACTCGGGG	GGGGGGGGGG
•	CTGAGGTCT	G CCTCGTGAAC	AAGGTGTTG	C TGACTCATAC	CAGGCCTGA	TCGCCCCATC
	ATCCAGCCA	G AAAGTGAGG	AGCCACGGT	T GATGAGAGCT	TTGTTGTAG	TGGACCAGTT
	GGTGATTTT	G AACTTTTGC	TTGCCACGG	A ACGGTCTGCC	TTGTCGGGAA	A GATGCGTGAT
25	CTGATCCTT	C AACTCAGCA	A AAGTTCGAT	T TATTCAACA	A AGCCGCCGT(CCGTCAAGTC
	AGCGTAATG	C TCTGCCAGT	TTACAACCA	A TTAACCAAT	r CTGATTAGA	A AAACTCATCG
	AGCATCAAA	T GAAACTGCA	A TTTATTCAT	A TCAGGATTA	r CAATACCAT	A TTTTTGAAAA
	AGCCGTTTC	T GTAATGAAG	G AGAAAACTC	A CCGAGGCAG	r TCCATAGGA	r GGCAAGATCC
	TGGTATCGG	ET CTGCGATTC	C GACTCGTCC	A ACATCAATA	C AACCTATTA	A TTTCCCCTCG
30	TCAAAAAT	AA GGTTATCAA	G TGAGAAATC	CATGAGTG	A CGACTGAAT	C CGGTGAGAAT
	GGCAAAAG(CT TATGCATTT	C TTTCCAGAC	T TGTTCAACA	G GCCAGCCAT	T ACGCTCGTCA
	TCAAAATC	AC TCGCATCAA	C CAAACCGT	TA TTCATTCGT	G ATTGCGCCT	G AGCGAGACGA
	AATACGCG	AT CGCTGTTAA	A AGGACAAT	ra caaacagga	A TCGAATGCA	A CCGGCGCAGG
	AACACTGC	CA GCGCATCAA	C AATATTTT	CA CCTGAATCA	G GATATTCTT	C TAATACCTGG

AATGCTGTTT TCCCGGGGAT CGCAGTGGTG AGTAACCATG CATCATGAG AGTACGGATA
AAATGCTTGA TGGTCGGAAG AGGCATAAAT TCCGTCAGCC AGTTTAGTCT GACCATCTCA
TCTGTAACAT CATTGGCAAC GCTACCTTTG CCATGTTCA GAAACAACTC TGGCGCATCG
GGCTTCCCAT ACAATCGATA GATTGTCGCA CCTGATTGCC CGACATTATC GCGAGCCCAT
TTATACCCAT ATAAATCAGC ATCCATGTTG GAATTTAATC GCGGCCTCGA GCAAGACGTT
TCCCGTTGAA TATGGCTCAT AACACCCCTT GTATTACTGT TTATGTAAGC AGACAGTTTT
ATTGTTCATG ATGATATATT TTTATCTTGT GCAATGTAAC ATCAGAGATT TTGAGACACA
ACGTGGCTTT CCCCCCCC CCATTATTGA AGCACTTTATC AGGGTTATTG TCTCATGAGC
GGATACATAT TTGAATGTAT TTAGAAAAAT AAACAAATAG GGGTTCCGCG CACATTTCCC
CGAAAAGTGC CACCTGACGT CTAAGAAACC ATTATTATCA TGACATTAAC CTATAAAAAT
AGGCGTATCA CGAGGCCCTT TCGTC (SEQ ID NO:9).

5

10

15

20

25

30

VIR - Vaccine vector V1R was constructed to obtain a minimum-sized vaccine vector without unneeded DNA sequences, which still retained the overall optimized heterologous gene expression characteristics and high plasmid yields that V1J and V1Jns afford. It was determined that (1) regions within the pUC backbone comprising the E. coli origin of replication could be removed without affecting plasmid yield from bacteria; (2) the 3'-region of the kanr gene following the kanamycin open reading frame could be removed if a bacterial terminator was inserted in its place; and, (3) ~300 bp from the 3'- half of the BGH terminator could be removed without affecting its regulatory function (following the original KpnI restriction enzyme site within the BGH element). V1R was constructed by using PCR to synthesize three segments of DNA from V1Jns representing the CMVintA promoter/BGH terminator, origin of replication, and kanamycin resistance elements, respectively. Restriction enzymes unique for each segment were added to each segment end using the PCR oligomers: SspI and XhoI for CMVintA/BGH; EcoRV and BamHI for the kan r gene; and, BclI and SalI for the ori r. These enzyme sites were chosen because they allow directional ligation of each of the PCR-derived DNA segments with subsequent loss of each site: EcoRV and SspI leave blunt-ended DNAs which are compatible for ligation while BamHI and BcII leave complementary overhangs as do Sall and Xhol. After obtaining these segments by PCR each segment was digested with the appropriate restriction enzymes indicated above and then ligated together in a single reaction mixture containing all three DNA segments. The 5'-end of the ori r was designed to include the T2 rho independent terminator sequence that is normally found in this region so that it could provide termination

20

25

30

information for the kanamycin resistance gene. The ligated product was confirmed by restriction enzyme digestion (>8 enzymes) as well as by DNA sequencing of the ligation junctions. DNA plasmid yields and heterologous expression using viral genes within V1R appear similar to V1Jns. The net reduction in vector size achieved was 1346 bp (V1Jns = 4.86 kb; V1R = 3.52 kb). PCR oligomer sequences used to 5 synthesize V1R (restriction enzyme sites are underlined and identified in brackets following sequence) are as follows: (1) 5'-GGTACAAATATTGGCTATTGG CCATTGCATACG-3' (SEQ ID NO:19) [SspI]; (2) 5'-CCACATCTCGAGGAAC CGGGTCAATTCTTCAGCACC-3' (SEQ ID NO:20) [XhoI] (for CMVintA/BGH segment); (3) 5'-GGTACAGATATCGGAAAGCCACGTTGTG TCTCAAAATC-3' 10 (SEQ ID NO:21) [EcoRV]; (4) 5'-CACATGGATCCGTAAT GCTCTGCCAGTGTT ACAACC-3' (SEQ ID NO:2) [BamHI], (for kanamycin resistance gene segment) (5) 5'-GGTACATG ATCACGTAGAAAAGATCA AAGGATCTTCTTG-3' (SEQ ID NO:23) [BclI]; (6) 5'-CCACATGTCGACCCGTAAA A'AGGCCGCGTTGCTGG-3' (SEQ ID NO:24): [SalI], (for E. coli origin of replication). 15

The nucleotide sequence of vector V1R is as follows:

TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG CTATTGGCCA TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT AATCAATTAC GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT ACATAACTTA CGGTAAATGG CCCGCCTGGC TGACCGCCCA ACGACCCCCG CCCATTGACG TCAATAATGA CGTATGTTCC CATAGTAACG CCAATAGGGA CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC TGCCCACTTG GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCCTAC TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG GTGATGCGGT TTTGGCAGTA CATCAATGGG CGTGGATAGC GGTTTGACTC ACGGGGATTT CCAAGTCTCC ACCCCATTGA CGTCAATGGG AGTTTGTTTT GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA CTCCGCCCCA TTGACGCAAA TGGGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG AGCTCGTTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACGCTGTT TTGACCTCCA TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA CGGTGCATTG GAACGCGGAT TCCCCGTGCC AAGAGTGACG TAAGTACCGC CTATAGAGTC TATAGGCCCA CCCCCTTGGC

	TTCTTATGCA	TGCTATACTG	TTTTTGGCTT	GGGGTCTATA	CACCCCGCT	TCCTCATGTT
	ATAGGTGATG	GTATAGCTTA	GCCTATAGGT	GTGGGTTATT	GACCATTATT	GACCACTCCC
	CTATTGGTGA	CGATACTTTC	CATTACTAAT	CCATAACATG	GCTCTTTGCC	ACAACTCTCT
	TTATTGGCTA	TATGCCAATA	CACTGTCCTT	CAGAGACTGA	CACGGACTCT	GTATTTTTAC
5	AGGATGGGGT	CTCATTTATT	ATTTACAAAT	TCACATATAC	AACACCACCG	TCCCCAGTGC
	CCGCAGTTTT	TATTAAACAT	AACGTGGGAT	CTCCACGCGA	ATCTCGGGTA	CGTGTTCCGG
	ACATGGGCTC	TTCTCCGGTA	GCGGCGGAGC	TTCTACATCC	GAGCCCTGCT	CCCATGCCTC
	CAGCGACTCA	TGGTCGCTCG	GCAGCTCCTT	GCTCCTAACA	GTGGAGGCCA	GACTTAGGCA
	CAGCACGATG	CCCACCACCA	CCAGTGTGCC	GCACAAGGCC	GTGGCGGTAG	GGTATGTGTC
10	TGAAAATGAG	CTCGGGGAGC	GGGCTTGCAC	CGCTGACGCA	TTTGGAAGAC	TTAAGGCAGC
	GGCAGAAGAA	GATGCAGGCA	GCTGAGTTGT	TGTGTTCTGA	TAAGAGTCAG	AGGTAACTCC
	CGTTGCGGTG	CTGTTAACGG	TGGAGGGCAG	TGTAGTCTGA	GCAGTACTCG	TTGCTGCCGC
	GCGCGCCACC	AGACATAATA	GCTGACAGAC	TAACAGACTG	TTCCTTTCCA	TGGGTCTTTT
	CTGCAGTCAC	CGTCCTTAGA	TCTGCTGTGC	CTTCTAGTTG	CCAGCCATCT	GTTGTTTGCC
15	CCTCCCCGT	GCCTTCCTTG	ACCCTGGAAG	GTGCCACTCC	CACTGTCCTT	TCCTAATAAA
	ATGAGGAAAT	TGCATCGCAT	TGTCTGAGTA	GGTGTCATTC	TATTCTGGGG	GGTGGGGTGG
	GGCAGCACAG	CAAGGGGGAG	GATTGGGAAG	ACAATAGCAG	GCATGCTGGG	GATGCGGTGG
	GCTCTATGGG	TACCCAGGTG	CTGAAGAATT	GACCCGGTTC	CTCCTGGGCC	AGAAAGAAGC
	AGGCACATCC	CCTTCTCTGT	GACACACCCT	GTCCACGCCC	CTGGTTCTTA	GTTCCAGCCC
20	CACTCATAGG	ACACTCATAG	CTCAGGAGGG	CTCCGCCTTC	AATCCCACCC	GCTAAAGTAC
	TTGGAGCGGT	CTCTCCCTCC	CTCATCAGCC	CACCAAACCA	AACCTAGCCT	CCAAGAGTGG
	GAAGAAATTA	AAGCAAGATA	GGCTATTAAG	TGCAGAGGGA	GAGAAAATGC	CTCCAACATG
	TGAGGAAGTA	ATGAGAGAAA	TCATAGAATT	TCTTCCGCTT	CCTCGCTCAC	TGACTCGCTG
	CGCTCGGTCG	TTCGGCTGCG	GCGAGCGGTA	TCAGCTCACT	CAAAGGCGGT	AATACGGTTA
25	TCCACAGAAT	CAGGGGATAA	CGCAGGAAAG	AACATGTGAG	CAAAAGGCCA	GCAAAAGGCC
	AGGAACCGTA	AAAAGGCCGC	GTTGCTGGCG	TTTTTCCATA	GGCTCCGCCC	CCCTGACGAG
	CATCACAAAA	ATCGACGCTC	AAGTCAGAGG	TGGCGAAACC	CGACAGGACT	ATAAAGATAC
	CAGGCGTTTC	CCCCTGGAAG	CTCCCTCGTG	CGCTCTCCTG	TTCCGACCCT	GCCGCTTACC
	GGATACCTGT	CCGCCTTTCT	CCCTTCGGGA	AGCGTGGCGC	TTTCTCAATG	CTCACGCTGT
30	AGGTATCTCA	GTTCGGTGTA	GGTCGTTCGC	TCCAAGCTGG	GCTGTGTGCA	CGAACCCCCC
	GTTCAGCCCG	ACCGCTGCGC	CTTATCCGGT	AACTATCGTC	TTGAGTCCAA	CCCGGTAAGA
	CACGACTTAT	CGCCACTGGC	AGCAGCCACT	GGTAACAGGA	TTAGCAGAGC	GAGGTATGTA
	GGCGGTGCTA	CAGAGTTCTT	GAAGTGGTGG	CCTAACTACG	GCTACACTAG	AAGGACAGTA
	TTTGGTATCT	GCGCTCTGCT	GAAGCCAGTT	ACCTTCGGAA	AAAGAGTTGG	TAGCTCTTGA

```
TCCGGCAAAC AAACCACCGC TGGTAGCGGT GGTTTTTTTG TTTGCAAGCA GCAGATTACG
    CGCAGAAAAA AAGGATCTCA AGAAGATCCT TTGATCTTTT CTACGGGGTC TGACGCTCAG
    TGGAACGAAA ACTCACGTTA AGGGATTTTG GTCATGAGAT TATCAAAAAG GATCTTCACC
    TAGATCCTTT TAAATTAAAA ATGAAGTTTT AAATCAATCT AAAGTATATA TGAGTAAACT
    TGGTCTGACA GTTACCAATG CTTAATCAGT GAGGCACCTA TCTCAGCGAT CTGTCTATTT
5
    CGTTCATCCA TAGTTGCCTG ACTCCGGGGG GGGGGGGGCGC TGAGGTCTGC CTCGTGAAGA
    AGGTGTTGCT GACTCATACC AGGCCTGAAT CGCCCCATCA TCCAGCCAGA AAGTGAGGGA
    GCCACGGTTG ATGAGAGCTT TGTTGTAGGT GGACCAGTTG GTGATTTTGA ACTTTTGCTT
    TGCCACGGAA CGGTCTGCGT TGTCGGGAAG ATGCGTGATC TGATCCTTCA ACTCAGCAAA
    AGTTCGATTT ATTCAACAAA GCCGCCGTCC CGTCAAGTCA GCGTAATGCT CTGCCAGTGT
10
    TACAACCAAT TAACCAATTC TGATTAGAAA AACTCATCGA GCATCAAATG AAACTGCAAT
    TTATTCATAT CAGGATTATC AATACCATAT TTTTGAAAAA GCCGTTTCTG TAATGAAGGA
    GAAAACTCAC CGAGGCAGTT CCATAGGATG GCAAGATCCT GGTATCGGTC TGCGATTCCG
     ACTCGTCCAA CATCAATACA ACCTATTAAT TTCCCCTCGT CAAAAATAAG GTTATCAAGT
     GAGAAATCAC CATGAGTGAC GACTGAATCC GGTGAGAATG GCAAAAGCTT ATGCATTTCT
15
     TTCCAGACTT GTTCAACAGG CCAGCCATTA CGCTCGTCAT CAAAATCACT CGCATCAACC
     AAACCGTTAT TCATTCGTGA TTGCGCCTGA GCGAGACGAA ATACGCGATC GCTGTTAAAA
     GGACAATTAC AAACAGGAAT CGAATGCAAC CGGCGCAGGA ACACTGCCAG CGCATCAACA
     ATATTTTCAC CTGAATCAGG ATATTCTTCT AATACCTGGA ATGCTGTTTT CCCGGGGATC
     GCAGTGGTGA GTAACCATGC ATCATCAGGA GTACGGATAA AATGCTTGAT GGTCGGAAGA
20
     GGCATAAATT CCGTCAGCCA GTTTAGTCTG ACCATCTCAT CTGTAACATC ATTGGCAACG
     CTACCTTTGC CATGTTTCAG AAACAACTCT GGCGCATCGG GCTTCCCATA CAATCGATAG
     ATTGTCGCAC CTGATTGCCC GACATTATCG CGAGCCCATT TATACCCATA TAAATCAGCA
     TCCATGTTGG AATTTAATCG CGGCCTCGAG CAAGACGTTT CCCGTTGAAT ATGGCTCATA
     ACACCCCTTG TATTACTGTT TATGTAAGCA GACAGTTTTA TTGTTCATGA TGATATATTT
25
     TTATCTTGTG CAATGTAACA TCAGAGATTT TGAGACACAA CGTGGCTTTC CCCCCCCCC
     CATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT
     TAGAAAAATA AACAAATAGG GGTTCCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC
     TAAGAAACCA TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC GAGGCCCTTT
     CGTC (SEQ ID NO:25).
30
```

EXAMPLE 2

Codon Optimized HIV-1 Pol and HIV-1 IA Pol Derivatives as DNA Vector Vaccines Synthesis of WT-optpol and IA-opt-pol Gene - Construction of both genes were conducted by Midland Certified Reagent Company (Midland, TX) following established strategies. Ten double stranded oligonucleotides, ranging from 159 to 340 bases long and encompassing the entire pol gene, were synthesized by solid state methods and cloned separately into pUC18. For the wt-pol gene, the fragments are as follows:

 $Bgl\Pi #1-Ecl$ 136 Π half site at 282 = pJS6A1-7PmII half site at #285 – Ecl136II half site at #597 10 = pJS6B2-5SspI half site at #600 - Ecl136II half site at #866 = pJS6C1-4SmaI half site at #869 – ApaI #1095 = pJS6D1-4ApaI #1095 - KpnI #1296 = pJS6E1-4KpnI #1296 - XcmI #1636 = pJS6F1-515 XcmI #1636 - NsiI #1847 = pJS6G1-2NsiI #1847 - BcII half site at #2174 = pJS6H1-14*Bcl*I half site at #2174 – *Sac*I #2333 = pJS6I1-2SacI #2333 - BglII #2577 = pJS6J1-1

5

20

25

30

EcoRI and HindIII sequences were added upstream of each 5' end and downstream of each 3' end, respectively, to allow cloning into the EcoRI-HindIII sites of pUC18.

The next stage of the synthesis was to consolidate these cassettes into three roughly equal fragments (alpha, beta, gamma) and was performed as follows:

Alpha: The SspI-HindIII small fragment of pJS6C1-4 was transferred into the Ecl136II-HindIII sites of pJS6B2-5 to give pJS6BC1-1. Into the EcoRI-PmII sites of this plasmid was inserted the EcoRI-Ecl136II small fragment of pJS6A1-7 to give pJS6 α 1-8.

Beta: The EcoRI-ApaI small fragment of pJS6D1-4 was inserted into the corresponding sites of pJS6E1-2 to give pJS6DE1-2. Also, the EcoRI-XcmI small fragment of pJS6F1-5 was inserted into the corresponding sites of pJS6G1-2 to give pJS6FG1-1. Then the EcoRI-KpnI small fragment of pJS6DE1-2 was inserted into the corresponding sites of pJS6FG1-1 to give pJS6 β 1-1.

Gamma: The SacI-HindIII small fragment of pJS6J1-1 was inserted into the corresponding sites of pJS6I1-2 to give pJS6IJ1-1. This plasmid was propagated through E. coli SCS110 (dam-/dcm-) to permit subsequent cleavage at the BcII site.

The BcII-HindIII small fragment of the unmethylated pJS6IJ1-1 was inserted into the BgIII-HindIII sites of pJS6H1-14 to give pJS6 χ 1-1.

The wt-pol alpha, beta, gamma were ligated into the entire sequence as follows:

The EcoRI-Ecl136II small fragment of pJS6α1-8 was inserted into the EcoRI-SmaI sites of pJS6β1-1 to give pJS6αβ2-1.

Into the NsiI-HindIII sites of this plasmid was inserted the NsiI-HindIII small fragment of pJS6 $\chi 1$ -1 to give pUC18-wt-pol. This final plasmid was completely resequenced in both strands.

To construct the entire IA-pol gene, only 3 new small fragments were synthesized:

PmlI half site at #285 - Ecl136II half site at #597 = pJS7B1-1

KpnI #1296 - XcmI #1636 = pJS7F1-2

NsiI #1847 - BgIII half site at #2174 = pJS7H1-5

These were then used in the same reconstruction strategy as described above to give pUC18-IA-pol.

20

25

30

Expression Vector Construction - pUC18-wt-pol and pUC18-IA-pol were digested with BgIII in order to isolate fragments containing the entire pol genes. V1R, V1Jns, V1Jns-tpa (Shiver, et al., 1995, Immune responses to HIV gp120 elicited by DNA vaccination. In Vaccines 95 (eds. Chanock, R. M., Brown, F., Ginsberg, H.S., & Norrby, E.) @ pp. 95-98; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; see also Example Section 1) were digested with BgIII. The cut vectors were then treated with calf intestinal alkaline phosphatase. Both wt-pol and

IA-pol genes were ligated into cut V1R using T4 DNA ligase (16 °C, overnight). Competent DH5α cells were transformed with aliquots of the ligation mixtures. Colonies were screened by restriction digestion of amplified plasmid isolates. Following a similar strategy, the *Bgl*Π fragment containing the IA-pol was subcloned into the *Bgl*Π site of V1Jns. To ligate the IA-pol gene into V1Jns-tpa, the IA-pol gene was PCR-amplified from V1R-IA-pol using pfu polymerase and the following pair of primers: 5'-GGTACAAGATCTCCGCCCCCATCTCCCCCATTGAGA-3' (SEQ ID NO:26), and 5'-CCACATAGATCTGCCCGGGCTTTAGTCCTCATC-3' (SEQ ID NO:27). The upstream primer was designed to remove the initiation met codon and place the pol gene in frame with the tpa leader coding sequence from V1Jns-tpa. The PCR product was purified from the agarose gel slab using Sigma

DNA Purification spin columns. The purified products were digested with $Bgl\Pi$ and subcloned into the $Bgl\Pi$ site of V1Ins-tpa.

Results - The codon humanized wt- and IA-pol genes were constructed via stepwise ligation of 10 synthetic dsDNA fragments (Ferretti, et al., 1986, *Proc. Natl. Acad. Sci. USA* 83: 599-603). For expression in mammalian systems, the IA-pol gene was subcloned into V1R, V1Jns, and V1Jns-tpa. All these vectors place the gene under the control of the human cytomegalovirus/intron A hybrid promoter (hCMVIA). The DNA sequence of the IA-pol gene and the expressed protein product are shown in Figure 2A-B. Subcloning into V1Jns-tpa attaches the leader sequence from human tissue-specific plasminogen activator (tpa) to the N-terminus of the IA-pol (Pennica, et al., 1983, *Nature* 301: 214-221) to allow secretion of the protein. The sequences of the tpa leader and the fusion junction are shown in Figure 3.

EXAMPLE 3

15

10

5

HIV-1 POL Vaccine - Rodent Studies

Materials - E. coli DH5α strain, penicillin, streptomycin, ACK lysis buffer, hepes, L-glutamine, RPMI1640, and ultrapure CsCl were obtained from Gibco/BRL (Grand Island, NY). Fetal bovine serum (FBS) was purchased from Hyclone. Kanamycin, Tween 20, bovine serum albumin, hydrogen peroxide (30%), 20 concentrated sulfuric acid, β-mercaptoethanol (β-ME), and concanavalin A were obtained from Sigma (St. Louis, MO). Female balb/c mice at 4-6 wks of age were obtained from Taconic Farms (Germantown, NY). 0.3-mL insulin syringes were purchased from Myoderm. 96-well flat bottomed Maxisorp plates were obtained form NUNC (Rochester, NY). HIV-1_{IIIB} RT p66 recombinant protein was obtained from 25 Advanced Biotechnologies, Inc. (Columbia, MD). 20-mer peptides were synthesized by Research Genetics (Huntsville, AL). Horseradish peroxidase (HRP)-conjugated rabbit anti-mouse IgG1 was obtained from ZYMED (San Francisco, CA). 1,2phenylenediamine dihydrochloride (OPD) tablets was obtained from DAKO (Norway). Purified rat anti-mouse IFN-gamma (IgG1, clone R4-6A2), biotin-30 conjugated rat anti-mouse IFN-gamma (IgG1, clone XMG 1.2), and strepavidinalkaline phosphatase conjugate were purchased from PharMingen (San Diego, CA). 1-STEP NBT/BCIP dye was obtained from Pierce Chemicals (Rockford, IL). 96-well Multiscreen membrane plate was purchased from Millipore (France). Cell strainer was obtained from Becton-Dickinson (Franklin Lakes, NJ).

Plasmid Preparation - E. coli DH5α cells expressing the pol plasmids were grown to saturation in LB broth supplemented with 100 ug/mL kanamycin. Plasmid were purified by standard CsCl method and solubilized in saline at concentrations greater than 5 mg/mL until further use.

Vaccination - The plasmids were prepared in phosphate-buffered saline and administered into balb/c by needle injection (28-1/2G insulin syringe) of 50 uL aliquot into each quad muscle. V1Jns-IApol was administered at 0.3, 3, 30 ug dose and for comparison, V1Jns-tpa-IApol was given at 30 ug dose. Immunizations were conducted at T=0 and T=8 wks (for select animals from the 30-ug dose cohorts).

5

10

15

20

25

30

ELISA Assay - At T=12 wks, blood samples were collected by making an incision of a tail vein and the serum separated. Anti-RT titers were obtained following standard secondary antibody-based ELISA. Briefly, Maxisorp plates were coated by overnight incubation with 100 uL of 1 ug/mL HIV-1 RT protein (in PBS). The plates were washed with PBS/0.05% Tween 20 and incubated for approx. 2h with 200 uL/well of blocking solution (PBS/0.05% tween/1% BSA). The blocking solution was decanted; 100 uL aliquot of serially diluted serum samples were added per well and incubated for 2 h at room temperature. The plates were washed and 100 uL of 1/1000-diluted HRP-rabbit anti-mouse IgG were added with 1 h incubation. The plates were washed thoroughly and soaked with 100 uL OPD/H₂O₂ solution for 15 min. The reaction was quenched by adding 100 uL of 0.5M H₂SO4 per well. OD₄₉₂ readings were recorded.

ELIspot - Spleens were collected from 5 mice/cohort at T=13-14 wks and pooled into a tube of 8-mL R10 medium (RPMI1640, 10% FBS, 2mM L-glutamine, 100U/mL Penicillin, 100 u/mL streptomycin, 10 mM Hepes, 50 uM β-ME). Multiscreen opaque plates were coated with 100μl/well of capture mAb (purified R4-6A2 diluted in PBS to 5μg/ml) at 4°C overnight. The plates were washed with PBS/Pen/Strep in hood and blocked with 200μl/well of complete R10 medium for 37°C for at least 2 hrs. The mouse spleens were ground on steel mesh, collected into 15ml tubes and centrifuged at 1200rpm for 10min. The pellet was treated in ACK buffer (4ml of lysis buffer per spleen) for 5min at room temperature to lyse red blood cells. The cell pellet was centrifuged as before, resuspended in K-medium (5ml per mouse spleen), filtered through a cell strainer and counted using a hemacytometer. Block medium was decanted from the plates and 100μl/well of cell samples (5.0x10e5 cells per well) plus antigens were added. Pol-specific CD4+ cells were stimulated

using a mixture of previously identified two epitope-containing peptides (aa641-660, aa731-750). Antigen-specific CD8+ cells were stimulated using a pool of four peptide epitope-containing peptides (aa201-220, aa311-330, aa571-590, aa781-800) or with individual peptides. A final concentration of 4 ug/mL per peptide was used.

5 Each splenocyte sample is tested for IFN-gamma secretion by adding the mitogen, concanavalin A. Plates were incubated at 37°C, 5% CO₂ for 20-24 h. The plates were washed with PBS/0.05% Tween 20 and soaked with 100 uL/well of 5 ug/mL biotin-conjugated rat anti-mouse IFN- mAb (clone XMG1.2) at 4°C overnight. The plates were washed and soaked with 100 uL/well 1/2500 dilution of strepavidin-AP (in PBS/0.005% Tween/5%FCS) for 30 min at 37 °C. Following a wash, spots were developed by incubating with 100µl/well 1-step NBT/BCIP for 6-10 min. The plates were washed with water and allowed to air dry. The number of spots in each wells were determined using a dissecting microscope and normalized to 10e6 cells.

Results - Single vaccination of balb/c mice with V1Jns-IApol is able to induce antigen-specific antibody (Figure 4) and T cell (Figure 5) responses in a dose response manner. IFN-gamma secretion from splenocytes can be detected from 3 and 30 ug cohort following stimulation with pools of peptides that contain CD4+ and CD8+ T cell epitopes. These epitopes were identified by (1) screening 20-mer peptides that encompass the entire pol sequence and overlap by 10 amino acid for ability to stimulate IFN-gamma secretion from vaccinee splenocytes, and (2) determining the T cell type (CD4+ or CD8+) by depleting either population in an Elispot assay. Addition of tpa leader sequence to the pol gene is able to induce comparable, if not slightly higher, frequencies of pol-specific CD4+ and CD8+ cells. A second immunization with either V1Jns-IApol and V1Jns-tpa-IApol resulted in effective boosting of the immune responses.

15

20

25

30

EXAMPLE 4

HIV-1 Pol Vaccine - Non Human Primate Studies

Materials - E. coli DH5α strain, penicillin, streptomycin, and ultrapure CsCl
were obtained from Gibco/BRL (Grand Island, NY). Kanamycin and
phytohemagluttinin (PHA-M) were obtained from Sigma (St. Louis, MO). 20-mer
peptides were synthesized by SynPep (Dublin, CA) and Research Genetics
(Huntsville, AL). 96-well Multiscreen Immobilon-P membrane plates were obtained
from Millipore (France). Strepavidin-alkaline phosphatase conjugate were purchased

5

10

15

20

25

30

form Pharmingen (San Diego, CA). 1-Step NBT/BCIP dye was obtained form Pierce Chemicals (Rockford, IL). Rat anti-human IFN-gamma mAb and biotin-conjugated anti-human IFN-gamma reagent were obtained from R&D Systems (Minneapolis, MN). Dynabeads M-450 anti-human CD4 were obtained from Dynal (Norway). HIVp24 antigen assay was purchased from Coulter Corporation (Miami, FL). HIV-1_{IIIB} RT p66 recombinant protein was obtained from Advanced Biotechnologies, Inc. (Columbia, MD). Plastic 8 well strips/plates, flat bottom, Maxisorp, are obtained from NUNC (Rochester, NY). HIV+ human serum 9711234 was obtained from Biological Specialty Corp.

Plasmid Preparation - E. coli DH5α cells expressing the pol plasmids were grown to saturation in LB supplemented with 100 ug/mL kanamycin. Plasmid were purified by standard CsCl method and solubilized in saline at concentrations greater than 5 mg/mL until further use.

Vaccination - Cohorts of 3 rhesus macaques (approx. 5-10 kg) were vaccinated with 5 mg dose of either V1Jns-IApol or V1Jns-tpa-IApol. The vaccine was administered by needle injection of two 0.5 mL aliquots of 5 mg/mL plasmid solution (in phosphate-buffered saline, pH 7.2) into both deltoid muscles. Prior to vaccination, the monkeys were chemically restraint with i.m. injection of 10 mg/kg ketamine. The animals were immunized 3x at 4 week intervals (T=0, 4, 8 wks).

Sample Collection - Blood samples were collected at $T=0,\,4,\,8,\,12,\,16,\,18$ wks; sera and PBMCs were isolated using established protocols.

ELIspot Assay - Immobilon-IP plates were coated with 100 uL/well of rat antihuman IFN-gamma mAb at 15 ug/mL at 4 °C overnight. The plates are then washed with PBS and block by adding 200 uL/well of R10 medium. 4x10e5 peripheral blood cells were plated per well and to each well, either media or one of the pol peptide pools (final concentration of 4 ug/mL per peptide) or PHA, a known mitogen, is added to a final volume of 100 uL. Duplicate wells were set up per sample per antigen and stimulation was performed for 20-24 h at 37 °C. The plates are then washed; biotinylated anti-human IFN-gamma reagent is added (0.1 ug/mL, 100 uL per well) and allowed to incubate for overnight at 4 °C. The plates are again washed and 100 uL of 1:2500 dilution of the strepavidin-alkaline phosphatase reagent (in PBS/0.005% Tween/5% FCS) is added and allowed to incubate for 2 h at ambient room temperature. After another wash, spots are developed by incubating with 100 uL/well of 1-step NBT/BCIP for 6-10 min. CD4- T cell depletion was performed by

adding 1 bead particle/10 cell of Dynabeads M450 anti-human CD4, prewashed with PBS, and incubating on the shaker at 4 °C for 30 min. The beads are fractionated magnetically and the unbound cells collected and quantified before plating onto the ELISpot assay plates (at 4x10e5 cells per well).

5

10

15

20

25

30

CTL Assay - Procedures for establishing bulk CTL culture with fresh or cryopreserved peripheral blood mononuclear cells (PBMC) are as follows. Twenty percent total PBMC were infected in 0.5 ml volume with recombinant vaccinia virus, Vac-tpaPol, respectively, at multiplicity of infection (moi) of 5 for 1 hr at 37°C, and then combined with the remaining PBMC sample. The cells were washed once in 10 ml R-10 medium, and plated in a 12 well plate at approximately 5 to 10 x 10⁶ cells/well in 4 ml R-10 medium. Recombinant human IL-7 was added to the culture at the concentration of 330 U/ml. Two or three days later, one milliliter of R-10 containing recombinant human IL-2 (100 U/ml) was added to each well. And twice weekly thereafter, two milliliters of cultured media were replaced with 2 ml fresh R-10 medium with rhIL-2 (100 U/ml). The lymphocytes were cultured at 37°C in the presence of 5% CO₂ for approximately 2 weeks, and used in cytotoxicity assay as described below. The effector cells harvested from bulk CTL cultures were tested against autologous B lymphoid cell lines (BLCL) sensitized with peptide pools. To prepare for the peptide-sensitized targets, the BLCL cells were washed once with R-10 medium, enumerated, and pulsed with peptide pool (about 4 to 8 ug/ml concentration for each individual peptide) in 1 ml volume overnight. A mock target was prepared by pulsing cells with peptide-free DMSO diluent to match the DMSO concentration in the peptide-pulsed targets. The cells were enumerated the next morning, and 1 x 10⁶ cells were resuspended in 0.5 ml R-10 medium. Five to ten microliters of Na 51 CrO₄ were added to the tubes at the same time, and the cells were incubated for 1 to 2 hr 37°C. The cells were then washed 3 times and resuspended at 5x10⁴ cells/ml in R-10 medium to be used as target cells. The cultured lymphocytes were plated with target cells at designated effector to target (E:T) ratios in triplicates in 96-well plates, and incubated at 37°C for 4 hours in the presence of 5% CO₂. A sample of 30 µl supernatant from each well of cell mixture was harvested onto a well of a Lumaplate-96 (Packard Instrument, Meriden, CT), and the plate was allowed to air dry overnight. The amount of ⁵¹Cr in the well was determined through betaparticle emission, using a plate counter from Packard Instrument. The percentage of specific lysis was calculated using the formula as: % specific lysis = (E-S) / (M-S).

5

10

15

20

25

30

The symbol E represents the average cpm released from target cells in the presence of effector cells, S is the spontaneous cpm released in the presence of medium only, and M is the maximum cpm released in the presence of 2% Triton X-100.

ELISA Assay - The pol-specific antibodies in the monkeys were measured in a competitive RT EIA assay, wherein sample activity is determined by the ability to block RT antigen from binding to coating antibody on the plate well. Briefly, Maxisorp plates were coated with saturating amounts of pol positive human serum (97111234). 250 uL of each sample is incubated with 15 uL of 266 ng/mL RT recombinant protein (in RCM 563, 1% BSA, 0.1% tween, 0.1% NaN₃) and 20 uL of lysis buffer (Coulter p24 antigen assay kit) for 15 min at room temperature. Similar mixtures are prepared using serially diluted samples of a standard and a negative control which defines maximum RT binding. 200 uL/well of each sample and standard were added to the washed plate and the plate incubated 16-24 h at room temperature. Bound RT is quantified following the procedures described in Coulter p24 assay kit and reported in milliMerck units per mL arbitrarily defined by the chosen standard.

Results - Repeated vaccinations with V1Jns-IApol induced in 1 of 3 monkeys (94R033) significant levels of antigen-specific T cell activation (Figure 6A-C and Table 2) and CTL killing of peptide-pulsed autologous cells (Figure 7A-B). A significant CD8+ component to the T cell responses in this animal was confirmed by peptide-stimulation of CD4-depleted PBMCs in an ELIspot assay (Table 2).

Immunization with V1Jns-tpa-IApol produced T cell responses from all 3 vaccinees (Figures 6A-C, Figure 7A-B; Table 2). Two (920078, 94R028) exhibited bulk CTL activity and detectable CD8+ components as measured by Elispot analyses of CD4-depleted PBMCs. For the third monkey (920073), the activated T cells were largely CD4+ (Table 2). Table 3 shows the time course data on the frequency of IFN-gamma secreting cells (SFC/million cells) upon antigen-specific stimulation for monkeys vaccinated 3x with either V1Jns-IApol or V1Jns-tpa-IApol (5 mg dose). At T=18 wks, CD4-cell depletion were performed; the reported values are the number of spots per million of fractionated cells and are not corrected for the resultant enrichment of CD8+ T cells. PBMCs were stimulated with peptide pools that represent either IA pol protein (mpol-1, mpol-2) or wt Pol (wtpol-1, wtpol-2).

TABLE 2

Vaccine	Animal No.	Antigen	T=0wk	T=4Wk	T=8Wk	T=12Wk	T=1	8 Wk
			Dose1	Dose 2	Dose 3			CD4-Dept
VIJns-IApol 5 mgs	9417.008	medum mpd-1 mpd-2 wipd-1 wipd-2	1 3 0	15 69 25 49 34	6 28 21 20 24	11 61 19 53 24	11 20 28 18 19	11 15 16
	94R013	medum mpd-1 mpd-2 wipd-1 wipd-2	0 0 1	14 9 15 9 6	6 63 24 50 21	9 25 36 33 29	18 34 24 18 25	11 9 15
	9417033	medum mpd-1 mpd-2 wlpd-1 wlpd-2	4 3 0	15 29 24 30 48	11 86 25 38 46	14 51 43 60 86	13 41 59 53 61	8 24 64
VIJns-tpa-lApd 5 mgs	920078	medum mpd-1 mpd-2 wtpd-1 wtpd-2	0 3 1	24 110 221 115 .218	13 120 130 53 204	11 119 551 70 490	14 155 289 116 194	11 11 145
	920073	medum mpd-1 mpd-2 wtpd-1 wtpd-2	0 0 0	13 36 29 20 25	3 51 16 35 16	15 113 83 100 79	15 90 115 74 61	6 14 34
	94IRO28	medum mpd-1 mpd-2 wlpd-1 wlpd-2	0 1 1	18 30 24 23 26	11 24 23 25 28	18 29 66 34 71	19 30 59 29 40	9 28 95
Nave	920072	medum mpd-1 mpd-2 wtpd-1 wtpd-2	1 0 1	19 24 24 18 23	3 11 5 13 14	38 25 28 20 33	9 4 6 6 14	4 6 5

5

10

For the Elispot assay, antigen specific stimulation were performed by using pools of 20-mer peptide pools based on the vaccine sequence. The vaccine pol sequence differs from the wild-type HIV-1 sequence by 9 point mutations, thereby affecting 16 of the 20-mer peptides in the pool. Comparable responses were observed in the vaccinees when these peptides are replaced with those using the wild-type sequences.

Four of the vaccinees gave anti-RT titers above background after 3 dosages of the plasmids (Table 2).

TABLE 3

Anti-RT levels in Rhesus Macaques Vaccinated 3x (4 week intervals) with 5 mgs of V1Jns-IApol or V1Jns-tpa-IApol expressed in mMU/mL.

Vacaine/Monkey	T=0Wk	T=4	T=8	T=12	T=16
V. C.	DOSE 1	DOSE 2	DOSE 3		
VIJns-IApol, 5 mg					
94R008	ND	<10	<10	15	14
94R013	· ND	<10	<10	<10_	<10
94R033	ND	<10	<10	25_	19
VIJns-tpa-lApal, 5 mg			<u> </u>		
920078	ND	<10	<10	35	17
920073	ND	<10	<10	<10	<10
94R028	ND	<10	<10	20	63
, ",	T			_	

15

20

EXAMPLE 5

Effect of Codon Optimization on In Vivo Expression and Cellular Immune Response of wt-pol

Materials and Methods - Extraction of virus-derived pol gene - The gene for RT-IN (wt-pol; a non-codon optimized wild type pol gene derived directly from the HIV IIIB genome) was extracted and amplified from the HIV IIIB genome using two primers, 5'-CAG GCG AGA TCT ACC ATG GCC CCC ATT AGC CCT ATT GAG ACT GTA-3' (SEQ ID NO:29) and 5'-CAG GCG AGA TCT GCC CGG GCT TTA ATC CTC ATC CTG TCT ACT TGC CAC-3' (SEQ ID NO:30), containing BglII sites.

The reaction contained 200 nmol of each primer, 2.5 U of pfu Turbo DNA polymerase (Stratagene, La Jolla, CA), 0.2 mM of each dNTPs, and the template DNA in 10mM KCl, 10mM (NH₄)₂SO₄, 20mM Tris-HCl pH 8.75, 2mM MgSO₄, 0.1% TritonX-100, 0.1mg/ml bovine serum albumin (BSA). Thermocycling

conditions were as follows: 20 cycles of 1 min at 95 °C, 1 min at 56 °C, and 4 mins at 72 °C with 15-min capping at 72 °C. The digested PCR fragment was subcloned into the *Bgl*II site of the expression plasmid V1Jns (Shiver, et al., 1995, Immune responses to HIV gp120 elicited by DNA vaccination. In Chanock, R. M., Brown, F., Ginsberg, H. S., and Norrby, E. (Eds.) *Vaccines 95*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp 95-98; see also Example section 1 herein) expression plasmid following similar procedures as described above. The ligation mixtures were then used to transform competent *E. coli* DH5 cells and screened by PCR amplification of individual colonies. Sequence of the entire gene insert was confirmed. All plasmid constructs for animal immunization were purified by CsCl method (Sambrook, et al., 1989, Fritsch and Maniatis, T. (Eds) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor).

5

10

15

20

25

30

In vitro expression in mammalian cells - 1.5x10⁶ 293 cells were transfected with 1 or 10 µg of V1R-wt-pol (codon optimized) and V1Jns-wt-pol (virus derived) using the Cell Phect kit and incubated for 48 h at 37 °C, 5% CO₂, 90% humidity. Supernatants and cell lysates were prepared and assayed for protein content using Pierce Protein Assay reagent (Rockford, IL). Aliquots containing equal amounts of total protein were loaded unto 10-20% Tris glycine gel (Novex, San Diego, CA) along with the appropriate molecular weight markers. The pol product was detected using anti-serum from a seropositive patient (Scripps Clinic, San Diego, CA) diluted 1:1000 and the bands developed using goat anti-human IgG-HRP (Bethyl, Montgomery, TX) at 1:2000 dilution and standard ECL reagent kit (Pharmacia LKB Biotechnology, Uppsala, Sweden).

Ultrasensitive RT activity assay of pol constructs - RT activities from codon optimized wt-pol and IA pol plasmids were analyzed by the Product-Enhanced Reverse Transcriptase (PERT) assay using Perkin Elmer 7700, Taqman technology (Arnold, et al., 1999, One-step fluorescent probe product-enhanced reverse transcriptase assay. In McClelland, M., Pardee, A. (Eds.) Expression genetics: accelerated and high-throughput methods. Biotechniques Books, Natick, MA, pp. 201-210). Background levels for this assay were determined using 1:100,000 dilution of lysates from mock (chemical treatment only, no vector) transfected 293 cells. This background range is set as RT/reaction tube of 0.00 to 56.28 which is taken from the mean value of 13.80 +/- 3 standard deviations (sd=14.16). Any individual value >56.28 would be considered positive for PERT assay. Cells lysates were prepared

similarly for the following samples: mock transfection with empty V1Jns vector; no vector control; transfection with V1Jns-tpa-pol (codon optimized); and transfection with V1Jns-IApol (codon optimized). Samples were serially diluted to 1:100,000 in PERT buffer and 24 replicates for each sample at this dilution were assayed for RT activity.

5

10

Rodent immunization with optimized and virus-derived pol plasmids - To compare the immunogenic properties of wt-pol (codon optimized) and virus-derived pol gene, cohorts of BALB/c mice (N=10) were vaccinated with 1 μg, 10 μg, and 100 μg doses of V1R-wt-pol (codon optimized) and V1Jns-wt-pol plasmid (virus derived). At 5 weeks post dose 1, 5 of 10 mice per cohort were boosted with the same dose of plasmid they initially received. In all cases, the vaccines were suspended or diluted in 6 mM sodium phosphate, 150 mM sodium chloride, pH 7.2, and the total dose was injected to both quadricep muscles in 50 μL aliquots using a 0.3-mL insulin syringe with 28-1/2G needles (Becton-Dickinson, Franklin Lakes, NJ).

Anti-RT ELISA - Anti-RT titers were obtained following standard secondary 15 antibody-based ELISA. Maxisorp plates (NUNC, Rochester, NY) were coated by overnight incubation with 100 µL of 1 µg /mL HIV-1 RT protein (Advanced Biotechnologies, Columbia, MD) in PBS. The plates were washed with PBS/0.05% Tween 20 using Titertek MAP instrument (Hunstville, AL) and incubated for approximately 2h with 200 µL/well of blocking solution (PBS/0.05% tween/1% 20 BSA). The blocking solution was decanted; 100 µL aliquot of serially diluted serum samples were added per well and incubated for 2 h at room temperature. An initial dilution of 100-fold is performed followed by 4-fold serial dilution. The plates were washed and 100 μ L of 1/1000-diluted HRP-rabbit anti-mouse IgG (ZYMED, San Francisco, CA) were added with 1 h incubation. The plates were washed thoroughly 25 and soaked with 100 µL 1,2-phenylenediamine dihydrochloride/hydrogen peroxide (DAKO, Norway) solution for 15 min. The reaction was quenched by adding 100 µL of 0.5M H₂SO4 per well. OD₄₉₂ readings were recorded using Titertek Multiskan MCC/340 with S20 stacker. Endpoint titers were defined as the highest serum dilution that resulted in an absorbance value of greater than or equal to 0.1 OD₄₉₂ (2.5 30 times the background value).

ELIspot assay - Antigen-specific INFγ-secreting cells from mouse spleens were detected using the ELIspot assay (Miyahira, et al., 1995, Quantification of antigen specific CD8⁺ T cells using an ELISPOT assay. *J. Immunol. Methods* 1995,

181, 45-54). Typically, spleens were collected from 3-5 mice/cohort and pooled into a tube of 8-mL complete RPMI media (RPMI1640, 10% FBS, 2mM L-glutamine, 100U/mL Penicillin, 100 u/mL streptomycin, 10 mM Hepes, 50 uM β-ME). Multiscreen opaque plates (Millipore, France) were coated with 100 µL/well of 5 5 µg/mL purified rat anti-mouse IFN-γ IgG1, clone R4-6A2 (Pharmingen, San Diego, CA), in PBS at 4°C overnight. The plates were washed with PBS/penicillin/streptomycin in hood and blocked with 200 µL/well of complete RPMI media for 37 °C for at least 2 h. The mouse spleens were ground on steel mesh, collected into 15ml tubes and centrifuged at 1200rpm for 10 min. The pellet 10 was treated with 4 mL ACK buffer (Gibco/BRL) for 5 min at room temperature to lyse red blood cells. The cell pellet was centrifuged as before, resuspended in complete RPMI media (5 ml per mouse spleen), filtered through a cell strainer and counted using a hemacytometer. Block media was decanted from the plates and to each well, $100 \mu L$ of cell samples (5×10^5 cells per well) and $100 \mu L$ of the antigen solution were added. To the control well, 100 µL of the media were added; for 15 specific responses, peptide pools containing either CD4⁺ or CD8⁺ epitopes were added. In all cases, a final concentration of 4 µg/mL per peptide was used. Each sample/antigen mixture were performed in triplicate wells. Plates were incubated at 37°C, 5% CO₂, 90% humidity for 20-24 h. The plates were washed with PBS/0.05% 20 Tween 20 and incubated with 100 µL/well of 1.25 µg/mL biotin-conjugated rat antimouse IFN-γ mAb, clone XMG1.2 (Pharmingen) at 4°C overnight. The plates were washed and incubated with 100 µL/well 1/2500 dilution of strepavidin-alkaline phosphatase conjugate (Pharmingen) in PBS/0.005% Tween/5% FBS for 30 min at 37 °C. Following a wash, spots were developed by incubating with 100 μl/well 1-step 25 NBT/BCIP (Pierce Chemicals) for 6-10 min. The plates were washed with water and allowed to air dry. The number of spots in each well was determined using a dissecting microscope and the data normalized to 10⁶ cell input.

Results - In vitro expression of Pol in mammalian cells - Heterologous expression of the optimized wt or IA pol genes (V1R-wt-pol (codon optimized), V1Jns-IApol (codon optimized), V1Jns-IApol (codon optimized), in 293 cells (Figure 8) yielded a single polypeptide of correct approximate molecular size (90-kDa) for the RT-IN fusion product. In contrast, no expression could be detected by transfecting cells with 1 and 10 μg of the V1Jns-wt-pol, which bears the virus-derived pol.

Ultrasensitive RT assay of cells transfected with Pol constructs - Table 4 summarizes the levels of polymerase activity from mock (vector only) control, IApol (codon optimized) and wt-pol plasmids (codon optimized). Results indicate that the wild-type POL transfected cells contained RT activity approximately 4-5 logs higher than the 293 cell only baseline values. Mock transfected cells contained activity no higher than baseline values. The RT activity from opt-IApol-transfected cells was also found to be no different than baseline values; no individual reaction tube resulted in RT activity higher than the established cut-off value of 56.

10

15

20

25

5

Table 4

Sample	Avg. RT/tube	Standard deviation	Minimum	Maximum
Vector only	16.25	18.52	0.0	42.99
IApol (codon	2.99	8.01	0.0	35.20
optimized)				
Wt-pol	126147	21338	68973	152007
(codon			Ì	
optimized)				

Comparative immunogenicity of optimized and virus-derived pol plasmid - To compare the *in vivo* potencies of both constructs, BALB/c mice (N=10 per group) were vaccinated with escalating doses (1, 10, 100 μg) of either V1Jns-wt-pol (virus derived) or V1R-wt-pol (codon optimized). At 5 wks post dose 1, 5 of 10 animals were randomly boosted with the same vaccine and dose they received initially. Figure 9 shows the geometric mean titers of the BALB/c cohorts determined at 2 wks past boost. No significant anti-RT titers can be observed from animals immunized with one or two doses of the wt-pol plasmid (virus derived). In contrast, animals vaccinated with the humanized gene construct gave cohort anti-RT titers (>1000) significantly above background levels at doses above 10 ug. The responses seen at 10 and 100 ug dose of V1R-wt-pol (codon optimized) were boosted approximately 10-fold with a second immunization, reaching titers as high as 10⁶. Spleens from all mice in each of the cohorts were collected to be analyzed for IFN-γ secretion following stimulation with mixtures of either CD4+ peptide epitopes or

CD8+ peptide epitopes. The results are shown in Figure 10. All wt-pol vaccinees did

not show any significant cellular response above the background controls. In contrast, strong antigen-stimulated IFN- γ secretion were observed in a dose-responsive manner from animals vaccinated with one or two doses of 10 or more μg of the wt-pol (codon optimized) construct.

The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are intended to fall within the scope of the appended claims.

10

WHAT IS CLAIMED IS:

- 1. A pharmaceutically acceptable DNA vaccine composition, which comprises:
 - (a) a DNA expression vector; and,
- 5 (b) a DNA molecule containing a codon optimized open reading frame encoding a Pol protein or inactivated Pol derivative thereof, wherein upon administration of the DNA vaccine to a host the Pol protein or inactivated Pol derivative is expressed and generates a cellular immune response against HIV-1 infection.

10

- 2. The DNA vaccine of claim 1 wherein the DNA molecule encodes wild type Pol.
- 3. The DNA vaccine of claim 2 wherein the DNA molecule comprises the nucleotide sequence as set forth in SEQ ID NO:1.
 - The DNA vaccine of claim 3 which is V1Jns-wt-pol.
- The DNA vaccine of claim 1 wherein the DNA molecule encodes an
 inactivated Pol derivative which contains a nucleotide sequence encoding a human
 tissue plasminogen activator leader peptide.
 - 6. The DNA vaccine of claim 5 wherein the DNA molecule comprises the nucleotide sequence as set forth in SEQ ID NO:5

- 7. The DNA vaccine of claim 6 which is V1Jns-tPA-wt-pol.
- 8. The DNA vaccine of claim 1 wherein the inactivated Pol protein contains at least one amino acid modification within each region of the Pol protein responsible for reverse transcriptase activity, RNase H activity and integrase activity, such that the inactivated Pol protein shows no substantial reverse transcriptase activity, RNase H activity and integrase activity.

9. The DNA vaccine of claim 8 wherein the DNA molecule comprises the nucleotide sequence as set forth in SEQ ID NO:3

10. The DNA vaccine of claim 9 which is V1Jns-IAPol.

5

- 11. The DNA vaccine of claim 8 wherein the DNA molecule encodes an inactivated Pol derivative which contains a nucleotide sequence encoding a human tissue plasminogen activator leader peptide.
- 10 12. The DNA vaccine of claim 11 wherein the DNA molecule comprises the nucleotide sequence as set forth in SEQ ID NO:7.
 - 13. The DNA vaccine of claim 7 which is V1Jns-tPA-IAPol.
- 14. A method for inducing an immune response against infection or disease caused by virulent strains of HIV which comprises administering into the tissue of a mammalian host a pharmaceutically acceptable DNA vaccine composition which comprises a DNA expression vector and a DNA molecule containing a codon optimized open reading frame encoding a Pol protein or inactivated Pol derivative thereof, wherein upon administration of the DNA vaccine to the vertebrate host the Pol protein or inactivated Pol derivative is expressed and generates the immune response.
 - 15. The method of claim 16 wherein the mammalian host is a human.

- 16. The method of claim 17 wherein the DNA vaccine is selected from the group consisting of V1Jns-WTPol, V1Jns-tPA-WTPol, V1Jns-IAPol and V1Jns-tPA-IAPol.
- 30 17. A substantially purified protein which comprises an amino acid sequence selected from the group consisting of SEQ ID NO:4, SEQ ID NO:6, and SEQ ID NO:8.

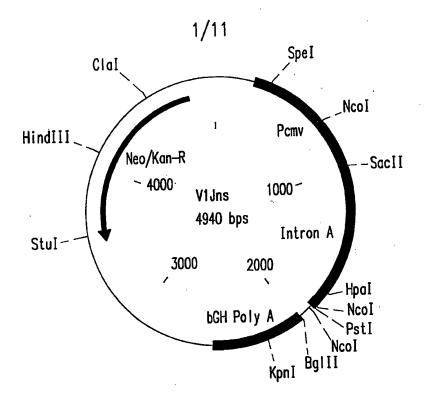


FIG.1A

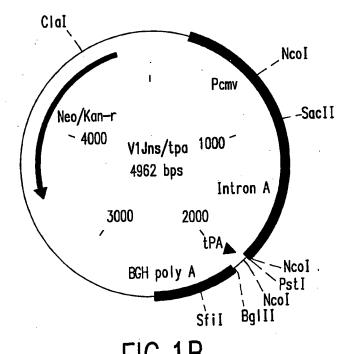


FIG.1B

SUBSTITUTE SHEET (RULE 26)

2/11

GCAGTGGCCCCTGACTGAGGAGAAGATCAAGGCCCTGGTGGAAATCTGCACTGAGATGGAGAAGGAGGGCCAAAATCTCCA sGInTrpProLeuThrGluCluLysIleLysAlaLeuVaIGluIleCysThrGluMetGluLysGluGlyLysIleSerL 30 40 50

AGATTGGCCCCGAGAACCCCTACAACACCCCTGTGTTTGCCATCAAGAAGAAGAAGACTCCACCAAGTGGAGGAAGCTGGTG
ysIteGtyProGtuAsnProTyrAsnThrProVatPheAtaIteLysLysLysAspSerThrLysTrpArgLysLeuVat
60 70

GACTTCAGGGAGCTGAACAAGAGGACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCACCCCGCTGGCCTGAAGAA AspPheArgGluLeuAsnLysArgThrGInAspPheTrpGluVaIGInLeuGlyIleProHisProAlaGlyLeuLysLy 80 90 100

GAAGAAGTCTGTGACTGTGCTGGCTGTGGGGGATGCCTACTTCTCTGTGCCCCTGGATGAGGACTTCAGGAAGTACACTG sLysLysSerValThrValLeuAlaValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrA 110 120 130

CCTTCACCATCCCCTCCATCAACAATGAGACCCCTGGCATCAGGTACCAGTACAATGTGCTGCCCCAGGGCTGGAAGGGC
IaPheThrlieProSerileAsnAsnGluThrProGlyIleArgTyrGInTyrAsnValLeuProGlnGlyTrpLysGly
140
150

TCCCCTGCCATCTTCCAGTCCTCCATGACCAAGATCCTGGAGCCCTTCAGGAAGCAGAACCCTGACATTGTGATCTACCA SerProAlaIlePheGInSerSerMetThrLysIleLeuGIuProPheArgLysGInAsnProAspIleVaIIleTyrGI 160 170 180

GTACATGCCTGCCCTGTATGTGGGCTCTGACCTGGAGATTGGGCAGCACAGGACCAAGATTGAGGAGCTGAGGCAGCACC
nTyrMetAlaAlaLeuTyrVaIGlySerAspLeuGluIleGlyGInHisArgThrLysIleGluGluLeuArgGInHisL
190 200 210

TGCTGAGGTGGGCCTGACCACCCCTGACAAGAAGCACCAGAAGGAGCCCCCCTTCCTGTGGATGGGCTATGAGCTGCAC euLeuArgTrpGlyLeuThrThrProAspLysLysHisGInLysGIuProProPheLeuTrpMetGIyTyrGIuLeuHis 220 230

CCCGACAAGTGGACTGTGCAGCCCATTGTGCTGCCTGAGAAGGACTCCTGGACTGTGAATGACATCCAGAAGCTGGTGGG ProAspLysTrpThrVaIGInProIIeVaILeuProGIuLysAspSerTrpThrVaIAsnAspIIeGInLysLeuVaIGI 240 250 260

CAAGCTGAACTGGGCCTCCCAAATCTACCCTGGCATCAAGGTGAGGCAGCTGTGCAAGCTGCTGAGGGGCACCAAGGCCC
yLysLeuAsnTrpAlaSerGinIleTyrProGlylleLysValArgGinLeuCysLysLeuLeuArgGlyThrLysAlaL
270 280 290

FIG.2A

SUBSTITUTE SHEET (RULE 26)

TGACTGAGGTGATCCCCCTGACTGAGGAGGCCTGAGCTGGAGCTGGAGCTGGAGAACAGGAGATCCTGAAGGAGCCTGTGCAT EuThrGluVollleProLeuThrGluGluAlaGluLeuGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHis 300 310

GGGGTGTACTATGACCCCTCCAAGGACCTGATTGCTGAGATCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAAATCTA GlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTy 320 330 340

CCAGGAGCCCTTCAAGAACCTGAAGACTGGCAAGTATGCCAGGATGAGGGGGGGCCCACACCAATGATGTGAAGCAGCTGA rGInGIuProPheLysAsnLeuLysThrGIyLysTyrAIaArgMetArgGIyAIaHisThrAsnAspVaILysGInLeuT 350 360 370

CTGAGGCTGTGCAGAAGATCACCACTGAGTCCATTGTGATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAG hrGiuAlaValGinLyslieThrThrGluSerlieVallieTrpGlyLysThrProLysPheLysLeuProlleGinLys 380 390

GGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCCATTGTGGGGGCCTGAGACCTTCTATGTGGCTGGGGCTGCCAACAGGG uValLysLeuTrpTyrGinLeuGluLysGluProlleValGlyAlaGluThrPheTyrValAlaGlyAlaAlaAsnArgG 430 440 450

AAGACTGCCCTCCAGGCCATCTACCTGGCCCTCCAGGACTCTGGCCTGGAGGTGAACATTGTGACTGCCTCCCAGTATGC LysThrAlaLeuGInAlaIleTyrLeuAlaLeuGInAspSerGlyLeuGluValAsnIleValThrAlaSerGInTyrAl 480 490 500

CCTGGGCATCATCCAGGCCCAGCCTGATCAGTCTGAGTCTGAGCTGGTGAACCAGATCATTGAGCAGCTGATCAAGAAGG aLeuGlyIIeIIeGInAlaGInProAspGInSerGluSerGluLeuValAsnGInIIeIIeGluGinLeuIleLysLysG 510 520

AGAAGGTGTACCTGGCCTGGCTGCCCACAAGGGCATTGGGGGCAATGAGCAGGTGGACAAGCTGGTGTCTGCTGGC
IuLysVaiTyrLeuAlaTrpVaiProAlaHisLysGlyIleGlyGlyAsnGluGInValAspLysLeuValSerAlaGly
540
550

ATCAGGAAGGTGCTGTTCCTGGATGGCATTGACAAGGCCCAGGATGAGCATGAGAAGTACCACTCCAACTGGAGGGCTAT IleArgLysValLeuPheLeuAspGlyIleAspLysAlaGlnAspGluHisGluLysTyrHisSerAsnTrpArgAlaMe 560 570 580

FIG.2B

SUBSTITUTE SHEET (RULE 26)

4/11

GGCCTCTGACTTCAACCTGCCCCCTGTGGTGGCTAAGGAGATTGTGGCCTCCTGTGACAAGTGCCAGCTGAAGGGGGAGG tAlaSerAspPheAsnLeuProProValValAlaLysGluIleValAlaSerCysAspLysCysGlnLeuLysGlyGluA 590 600 610

GCTGTGCATGTGGCCTCCGGCTACATTGAGGCTGAGGTGATCCCTGCTGAGACAGGCCAGGAGACTGCCTACTTCCTGCT AlaValHisValAlaSerGlyTyrIleGluAlaGluValIleProAlaGluThrGlyGlnGluThrAlaTyrPheLeuLe 640 650 660

GAAGCTGGCTGGCAGGTGGCCTGTGAAGACCATCCACACTGCCAATGGCTCCAACTTCACTGGGGCCACAGTGAGGGCTG
uLysLeuAlaGlyArgTrpProValLysThrIIeHisThrAlaAsnGlySerAsnPheThrGlyAlaThrValArgAlaA
670 680 690

CCTGCTGGTGGGCTGGCATCAAGCAGGAGTTTGGCATCCCCTACAACCCCCAGTCCCAGGGGGTGGTGGCCTCCATGAAC laCysTrpTrpAlaGlyIleLysGlnGluPheGlyIleProTyrAsnProGinSerGlnGlyValValAlaSerMetAsn 700 710

AAGGAGCTGAAGAAGATCATTGGGCAGGTGAGGGACCAGGCTGAGCACCTGAAGACAGCTGTGCAGATGGCTGTTCAT LysG1uLeuLysLysIIeIIeGIyGinVaIArgAspGinAlaGluHisLeuLysThrAlaVaIGInMetAlaVaIPheII 720 730 740

CCACAACTTCAAGAGGAAGGGGGGCATCGGGGGCTACTCCGCTGGGGAGAGGATTGTGGACATCATTGCCACAGACATCC eHisAsnPheLysArgLysGiyGiyIleGiyGiyTyrSerAloGiyGluArgIleValAspIleIleAlaThrAspIleG 750 760 770

AGACCAAGGAGCTCCAGAAGCAGATCACCAAGATCCAGAACTTCAGGGTGTACTACAGGGACTCCAGGAACCCCCTGTGG
InThrLysGluLeuGInLysGInIleThrLysIleGInAsnPheArgVaITyrTyrArgAspSerArgAsnProLeuTrp
780
790

AAGGGCCCTGCCAAGCTGCTGTGGAAGGGGGGGGGGGTGTGGTGATCCAGGACAACTCTGACATCAAGGTGGTGCCCAG LysGIyProAiaLysLeuLeuTrpLysGIyGIuGIyAIaVaIVaIIIeGInAspAsnSerAspIIeLysVaIVaIProAr 800 810 820

AAAGCCCGGGCAGATCT (SEQ ID NO: 3) $Xx = \overline{BgIII}$

FIG.2C

5/11

GATCACCATGGATGCAATGAAGAGGGCTCTGCTGTGTGCTGCTGTGTGGAGCAGTCTTCGTTTCGC MetAspAloMetLysArgGlyLeuCysCysVolLeuLeuLeuCysGlyAloVolPheVolSerP -25 RoSerGIUIIeSerAlaProlleGerProlleGluThrValProValLysLeuLysProGlyMetAspGIy (within SEQ ID NO: 7)

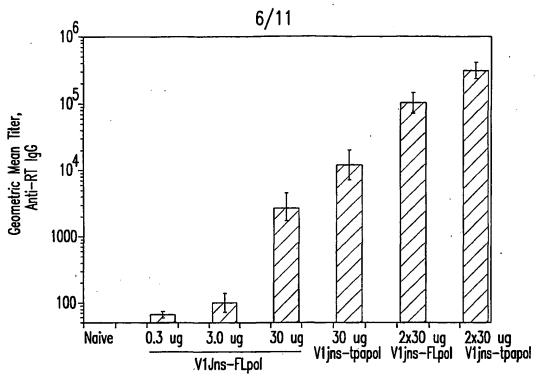
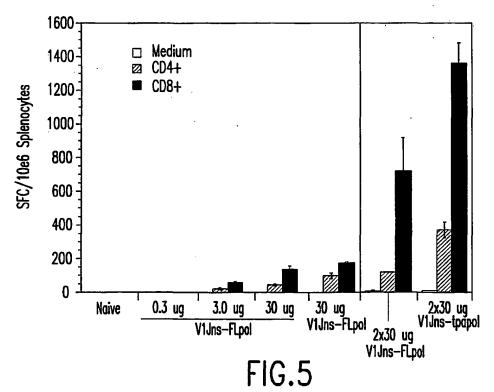


FIG.4



SUBSTITUTE SHEET (RULE 26)



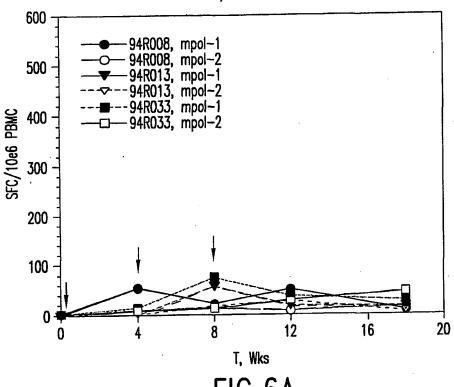


FIG.6A

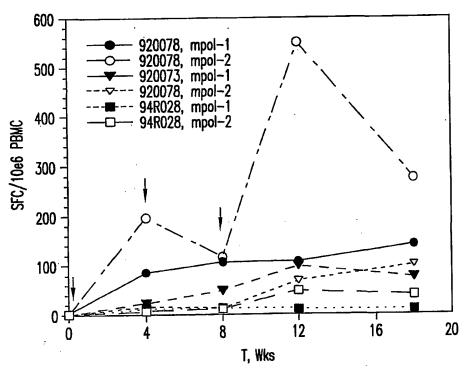
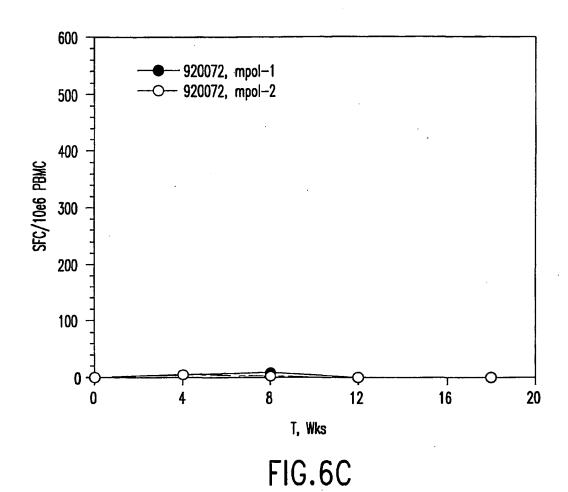
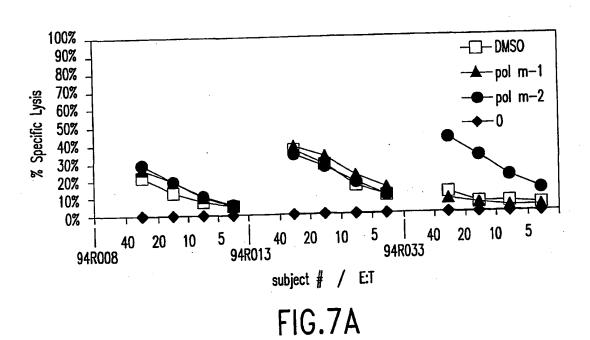


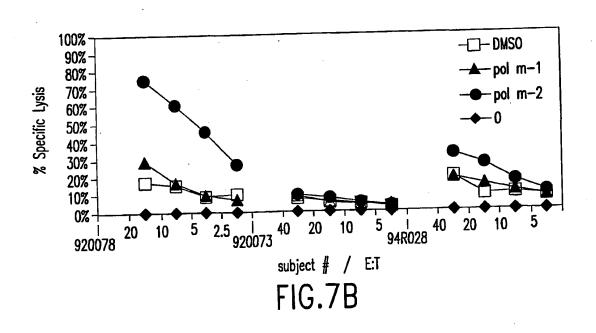
FIG.6B

SUBSTITUTE SHEET (RULE 26)



SUBSTITUTE SHEET (RULE 26)





SUBSTITUTE SHEET (RULE 26)

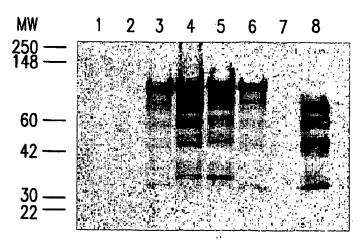
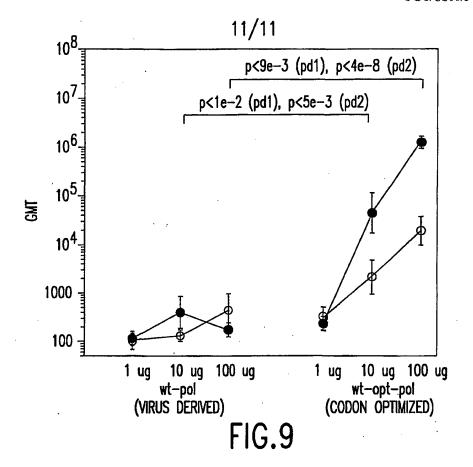


FIG.8



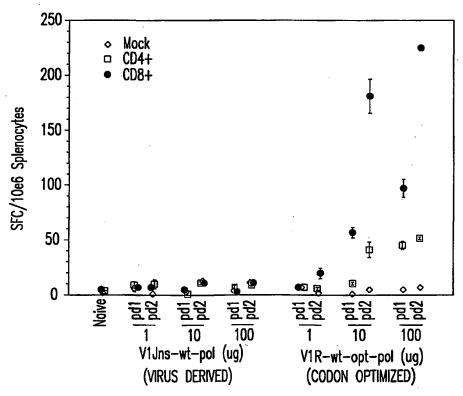


FIG. 10
SUBSTITUTE SHEET (RULE 26)

SEQUENCE LISTING

<110> Merck & Co., Inc.

130

<120> POLYNUCLEOTIDE VACCINES EXPRESSING CODON OPTIMIZED HIV-1 POL AND MODIFIED HIV-1 POL

<130> 20608Y PCT <160> 30 <170> FastSEQ for Windows Version 4.0 <210> 1 <211> 2577 <212> DNA <213> Human Immunodeficiency Virus-1 <220> <221> CDS <222> (10)...(2562) <400> 1 agatctacc atg gcc ccc atc tcc ccc att gag act gtg cct gtg aag ctg Met Ala Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu aag cct ggc atg gat ggc ccc aag gtg aag cag tgg ccc ctg act gag Lys Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu 99 gag aag atc aag gcc ctg gtg gaa atc tgc act gag atg gag aag gag Glu Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu 147 ggc aaa atc tcc aag att ggc ccc gag aac ccc tac aac acc cct gtg Gly Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val 195 ttt gcc atc aag aag aag gac tcc acc aag tgg agg aag ctg gtg gac 243 Phe Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp ttc agg gag ctg aac aag agg acc cag gac ttc tgg gag gtg cag ctg 291 Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu ggc atc ccc cac ccc gct ggc ctg aag aag aag tct gtg act gtg 339 Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val 100 ctg gat gtg ggg gat gcc tac ttc tct gtg ccc ctg gat gag gac ttc 387 Leu Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe 115 agg aag tac act gcc ttc acc atc ccc tcc atc aac aat gag acc cct 435 Arg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro

135

ggc Gly	atc Ile	agg Arg 145	tac Tyr	cag Gln	tac Tyr	aat Asn	gtg Val 150	ctg Leu	ccc Pro	cag Gln	ggc Gly	tgg Trp 155	aag Lys	ggc Gly	tcc Ser	483
cct Pro	gcc Ala 160	atc Ile	ttc Phe	cag Gln	tcc Ser	tcc Ser 165	atg Met	acc Thr	aag Lys	atc Ile	ctg Leu 170	gag Glu	ccc Pro	ttc Phe	agg Arg	531
aag Lys 175	cag Gln	aac Asn	cct Pro	gac Asp	att Ile 180	gtg Val	atc Ile	tac Tyr	cag Gln	tac Tyr 185	atg Met	gat Asp	gac Asp	ctg Leu	tat Tyr 190	579
gtg Val	ggc Gly	tct Ser	Asp	ctg Leu 195	gag Glu	att Ile	GJA aaa	cag Gln	cac His 200	agg Arg	acc Thr	aag Lys	att Ile	gag Glu 205	gag Glu	627
ctg Leu	agg Arg	cag Gln	cac His 210	ctg Leu	ctg Leu	agg Arg	tgg Trp	ggc Gly 215	ctg Leu	acc Thr	acc Thr	cct Pro	gac Asp 220	aag Lys	aag Lys	675
cac His	cag Gln	aag Lys 225	gag Glu	ccc Pro	ccc Pro	ttc Phe	ctg Leu 230	tgg Trp	atg Met	ggc Gly	tat Tyr	gag Glu 235	ctg Leu	cac His	ccc Pro	723
gac Asp	aag Lys 240	tgg Trp	act Thr	gtg Val	cag Gln	ccc Pro 245	att Ile	gtg Val	ctg Leu	cct Pro	gag Glu 250	aag Lys	gac Asp	tcc Ser	tgg Trp	771
act Thr 255	gtg Val	aat Asn	gac Asp	atc Ile	cag Gln 260	aag Lys	ctg Leu	gtg Val	ggc	aag Lys 265	ctg Leu	aac Asn	tgg Trp	gcc Ala	tcc Ser 270	819
caa Gln	atc Ile	tac Tyr	cct Pro	ggc Gly 275	atc Ile	aag Lys	gtg Val	agg Arg	cag Gln 280	ctg Leu	tgc Cys	aag Lys	ctg Leu	ctg Leu 285	agg Arg	867
ggc Gly	acc Thr	aag Lys	gcc Ala 290	ctg Leu	act Thr	gag Glu	gtg Val	atc Ile 295	ccc Pro	ctg Leu	act Thr	gag Glu	gag Glu 300	gct Ala	gag Glu	915
ctg Leu	gag Glu	ctg Leu 305	gct Ala	gag Glu	aac Asn	agg Arg	gag Glu 310	atc Ile	ctg Leu	aag Lys	gag Glu	cct Pro 315	gtg Val	cat His	Gly ggg	963
gtg Val	tac Tyr 320	tat Tyr	gac Asp	ccc Pro	tcc Ser	aag Lys 325	gac Asp	ctg Leu	att Ile	gct Ala	gag Glu 330	atc Ile	cag Gln	aag Lys	cag Gln	1011
ggc Gly 335	cag Gln	ggc Gly	cag Gln	tgg Trp	acc Thr 340	tac Tyr	caa Gln	atc Ile	tac Tyr	cag Gln 345	gag Glu	ccc Pro	ttc Phe	aag Lys	aac Asn 350	1059
ctg Leu	aag Lys	act Thr	ggc Gly	aag Lys 355	tat Tyr	gcc Ala	agg Arg	atg Met	agg Arg 360	ggg Gly	gcc Ala	cac His	acc Thr	aat Asn 365	gat Asp	1107
gtg Val	aag Lys	cag Gln	ctg Leu 370	act Thr	gag Glu	gct Ala	gtg Val	cag Gln 375	aag Lys	atc Ile	acc Thr	act Thr	gag Glu 380	tcc Ser	att Ile	1155

gtg Val	atc Ile	tgg Trp 385	ggc	aag Lys	acc Thr	ccc Pro	aag Lys 390	ttc Phe	aag Lys	ctg Leu	ccc Pro	atc Ile 395	cag Gln	aag Lys	gag Glu	1203
acc Thr	tgg Trp 400	gag Glu	acc Thr	tgg Trp	tgg Trp	act Thr 405	gag Glu	tac Tyr	tgg Trp	cag Gln	gcc Ala 410	acc Thr	tgg Trp	atc Ile	cct Pro	1251
gag Glu 415	tgg Trp	gag Glu	ttt Phe	gtg Val	aac Asn 420	acc Thr	ccc Pro	ccc Pro	ctg Leu	gtg Val 425	aag Lys	ctg Leu	tgg Trp	tac Tyr	cag Gln 430	1299
								gct Ala								1347
gct Ala	gcc Ala	aac Asn	agg Arg 450	gag Glu	acc Thr	aag Lys	ctg Leu	ggc Gly 455	aag Lys	gct Ala	ggc Gly	tat Tyr	gtg Val 460	acc Thr	aac Asn	1395
agg Arg	ggc Gly	agg Arg 465	cag Gln	aag Lys	gtg Val	gtg Val	acc Thr 470	ctg Leu	act Thr	gac Asp	acc Thr	acc Thr 475	aac Asn	cag Gln	aag Lys	1443
act Thr	gag Glu 480	ctc Leu	cag Gln	gcc Ala	atc Ile	tac Tyr 485	ctg Leu	gcc Ala	ctc Leu	cag Gln	gac Asp 490	tct Ser	ggc Gly	ctg Leu	gag Glu ··	1491
gtg Val 495	aac Asn	att Ile	gtg Val	act Thr	gac Asp 500	tcc Ser	cag Gln	tat Tyr	gcc Ala	ctg Leu 505	ggc Gly	atc Ile	atc Ile	cag Gln	gcc Ala 510	1539
cag Gln	cct Pro	gat Asp	cag Gln	tct Ser 515	gag Glu	tct Ser	gag Glu	ctg Leu	gtg Val 520	aac Asn	cag Gln	atc Ile	att Ile	gag Glu 525	cag Gln	1587
ctg Leu	atc Ile	aag Lys	aag Lys 530	gag Glu	aag Lys	gtg Val	tac Tyr	ctg Leu 535	gcc Ala	tgg Trp	gtg Val	cct Pro	gcc Ala 540	cac His	aag Lys	1635
ggc Gly	att Ile	ggg Gly 545	ggc Gly	aat Asn	gag Glu	cag Gln	gtg Val 550	gac Asp	aag Lys	ctg Leu	gtg Val	tct Ser 555	gct Ala	ggc Gly	atc Ile	1683
agg Arg	aag Lys 560	gtg Val	ctg Leu	ttc Phe	ctg Leu	gat Asp 565	ggc Gly	att Ile	gac Asp	aag Lys	gcc Ala 570	cag Gln	gat Asp	gag Glu	cat His	1731
gag Glu 575	aag Lys	tac Tyr	cac His	tcc Ser	aac Asn 580	tgg Trp	agg Arg	gct Ala	atg Met	gcc Ala 585	tct Ser	gac Asp	ttc Phe	aac Asn	ctg Leu 590	1779
ccc Pro	cct Pro	gtg Val	gtg Val	gct Ala 595	aag Lys	gag Glu	att Ile	gtg Val	gcc Ala 600	tcc Ser	tgt Cys	gac Asp	aag Lys	tgc Cys 605	cag Gln	1827
ctg Leu	aag Lys	Gly aaa	gag Glu 610	gcc Ala	atg Met	cat His	gly ggg	cag Gln 615	gtg Val	gac Asp	tgc Cys	tcc Ser	cct Pro 620	ggc Gly	atc Ile	1875

tgg Trp	cag Gln	ctg Leu 625	gac Asp	tgc Cys	acc Thr	cac His	ctg Leu 630	gag Glu	ggc Gly	aag Lys	gtg Val	atc Ile 635	ctg Leu	gtg Val	gct Ala	1923
gtg Val	cat His 640	gtg Val	gcc Ala	tcc Ser	ggc Gly	tac Tyr 645	att Ile	gag Glu	gct Ala	gag Glu	gtg Val 650	atc Ile	cct Pro	gct Ala	gag Glu	1971
aca Thr 655	ggc Gly	cag Gln	gag Glu	act Thr	gcc Ala 660	tac Tyr	ttc Phe	ctg Leu	ctg Leu	aag Lys 665	ctg Leu	gct Ala	ggc Gly	agg Arg	tgg Trp 670	2019
cct Pro	gtg Val	aag Lys	acc Thr	atc Ile 675	cac His	act Thr	gac Asp	aat Asn	ggc 680	tcc Ser	aac Asn	ttc Phe	act Thr	ggg Gly 685	gcc Ala	2067
aca Thr	gtg Val	agg Arg	gct Ala 690	gcc Ala	tgc Cys	tgg Trp	tgg Trp	gct Ala 695	ggc Gly	atc Ile	aag Lys	cag Gln	gag Glu 700	ttt Phe	Gly ggc	2115
atc Ile	ccc Pro	tac Tyr 705	aac Asn	ccc Pro	cag Gln	tcc Ser	cag Gln 710	ggg	gtg Val	gtg Val	gag Glu	tcc Ser 715	atg Met	aac Asn	aag Lys	2163
gag Glu	ctg Leu 720	aag Lys	aag Lys	atc Ile	att Ile	ggg Gly 725	cag Gln	gtg Val	agg Arg	gac Asp	cag Gln 730	gct Ala	gag Glu	cac His	·ctg Leu	2211
aag Lys 735	aca Thr	gct Ala	gtg Val	cag Gln	atg Met 740	gct Ala	gtg Val	ttc Phe	atc Ile	cac His 745	aac Asn	ttc Phe	aag Lys	agg Arg	aag Lys 750	2259
Gly ggg	ggc Gly	atc Ile	Gly	ggc Gly 755	tac Tyr	tcc Ser	gct Ala	Gly	gag Glu 760	agg Arg	att Ile	gtg Val	gac Asp	atc Ile 765	att Ile	2307
gcc Ala	aca Thr	gac Asp	atc Ile 770	cag Gln	acc Thr	aag Lys	gag Glu	ctc Leu 775	cag Gln	aag Lys	cag Gln	atc Ile	acc Thr 780	aag Lys	atc Ile	2355
cag Gln	aac Asn	ttc Phe 785	agg Arg	gtg Val	tac Tyr	tac Tyr	agg Arg 790	gac Asp	tcc Ser	agg Arg	aac Asn	ccc Pro 795	ctg Leu	tgg Trp	aag Lys	2403
ggc Gly	cct Pro 800	gcc Ala	aag Lys	ctg Leu	ctg Leu	tgg Trp 805	aag Lys	ggg Gly	gag Glu	Gly	gct Ala 810	gtg Val	gtg Val	atc Ile	cag Gln	2451
gac Asp 815	aac Asn	tct Ser	gac Asp	atc Ile	aag Lys 820	gtg Val	gtg Val	ccc Pro	agg Arg	agg Arg 825	Lys	gcc Ala	aag Lys	atc Ile	atc Ile 830	2499
agg Arg	gac Asp	tat Tyr	ggc Gly	aag Lys 835	cag Gln	atg Met	gct Ala	Gly	gat Asp 840	gac Asp	tgt Cys	gtg Val	gcc Ala	tcc Ser 845	agg Arg	2547
			gac Asp 850	*	agc	ccgg	gca	gatc	t	·						2577

<211> 850 <212> PRT <213> Human Immunodeficiency Virus-1 <400> 2 Met Ala Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys 2.0 Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala 50 55 60 Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg 65 75 75 80 Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu Asp 100 105 Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln
165 170 175 Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val Gly 180 185 Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys His Gln 210 215 220 Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln 325 330 335 Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Asp Gly Ala Ala

```
Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly
                        455
                                            460
Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Glu
                  470
                                        475
Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn
               485
                                    490
                                                        495
Ile Val Thr Asp Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro
           500
                               505
                                                    510
Asp Gln Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile 515 520 525
Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile
   530
                                           540
                       535
Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys 545 550 560
               550
Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys
565 570 575
               565
                                  570
Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro
           580
                               585
                                                  590
Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys
        595
                           600
                                                605
Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln
                       615
                                           620
Leu Asp Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His
                 630
                                        635
Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly
645 650 655
                                   650
Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val
                                                    670
           660
                               665
Lys Thr Ile His Thr Asp Asn Gly Ser Asn Phe Thr Gly Ala Thr Val
                           680
                                               685
     675
Arg Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro
    690
                       695
                                           700
Tyr Asn Pro Gln Ser Gln Gly Val Val Glu Ser Met Asn Lys Glu Leu
                   710
                                        715
Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr 725 730 735
Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly 740 745 750
Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr
        755
                            760
Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn
                        775
                                            780
Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro
785
                   790
                                        795
Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn 815
               805
                                    810
Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp
           820
                               825
                                                    830
Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp
                            840
Glu Asp
   850
<210> 3
<211> 2577
<212> DNA
<213> Human Immunodeficiency Virus-1
<220>
<221> CDS
<222> (10)...(2562)
```

<400 agat			et Al								ar Va				ag ctg /s Leu	51
								gtg Val								99
gag Glu	aag Lys	atc Ile	aag Lys	gcc Ala 35	ctg Leu	gtg Val	gaa Glu	atc Ile	tgc Cys 40	act Thr	gag Glu	atg Met	gag Glu	aag Lys 45	gag Glu	147
								gag Glu 55								195
								acc Thr								243
								cag Gln								291
															gtg Val 110	339
								tct Ser								387
agg Arg	aag Lys	tac Tyr	act Thr 130	gcc Ala	ttc Phe	acc Thr	atc Ile	ccc Pro 135	tcc Ser	atc Ile	aac Asn	aat Asn	gag Glu 140	acc Thr	cct Pro	435
ggc Gly	atc Ile	agg Arg 145	tac Tyr	cag Gln	tac Tyr	aat Asn	gtg Val 150	ctg Leu	ccc Pro	cag Gln	ggc Gly	tgg Trp 155	aag Lys	ggc Gly	tcc Ser	483
cct Pro	gcc Ala 160	atc Ile	ttc Phe	cag Gln	tcc Ser	tcc Ser 165	atg Met	acc Thr	aag Lys	atc Ile	ctg Leu 170	gag Glu	ccc Pro	ttc Phe	agg Arg	531
aag Lys 175	cag Gln	aac Asn	cct Pro	gac Asp	att Ile 180	gtg Val	atc Ile	tac Tyr	cag Gln	tac Tyr 185	atg Met	gct Ala	gcc Ala	ctg Leu	tat Tyr 190	579
gtg Val	ggc Gly	tct Ser	gac Asp	ctg Leu 195	gag Glu	att Ile	GJÀ aaa	cag Gln	cac His 200	agg Arg	acc Thr	aag Lys	att Ile	gag Glu 205	gag Glu	627
ctg Leu	agg Arg	cag Gln	cac His 210	ctg Leu	ctg Leu	agg Arg	tgg Trp	ggc Gly 215	ctg Leu	acc Thr	acc Thr	cct Pro	gac Asp 220	aag Lys	aag Lys	675
cac His	cag Gln	aag Lys 225	gag Glu	ccc Pro	ccc Pro	ttc Phe	ctg Leu 230	tgg Trp	atg Met	ggc Gly	tat Tyr	gag Glu 235	ctg Leu	cac His	ccc Pro	723

gac Asp	aag Lys 240	tgg Trp	act Thr	gtg Val	cag Gln	ccc Pro 245	att Ile	gtg Val	ctg Leu	cct Pro	gag Glu 250	aag Lys	gac Asp	tcc Ser	tgg Trp	771
act Thr 255	gtg Val	aat Asn	gac Asp	atc Ile	cag Gln 260	aag Lys	ctg Leu	gtg Val	ggc Gly	aag Lys 265	ctg Leu	aac Asn	tgg Trp	gcc Ala	tcc Ser 270	819
caá Gln	atc Ile	tac Tyr	cct Pro	ggc Gly 275	atc Ile	aag Lys	gtg Val	agg Arg	cag Gln 280	ctg Leu	tgc Cys	aag Lys	ctg Leu	ctg Leu 285	agg Arg	867
ggc Gly	acc Thr	aag Lys	gcc Ala 290	ctg Leu	act Thr	gag Glu	gtg Val	atc Ile 295	ccc Pro	ctg Leu	act Thr	gag Glu	gag Glu 300	gct Ala	gag Glu	915
ctg Leu	gag Glu	ctg Leu 305	gct Ala	gag Glu	aac Asn	agg Arg	gag Glu 310	atc Ile	ctg Leu	aag Lys	gag Glu	cct Pro 315	gtg Val	cat His	Gly ggg	963
gtg Val	tac Tyr 320	tat Tyr	gac Asp	ccc Pro	tcc Ser	aag Lys 325	gac Asp	ctg Leu	att Ile	gct Ala	gag Glu 330	atc Ile	cag Gln	aag Lys	cag Gln	1011
ggc Gly 335	cag Gln	ggc Gly	cag Gln	tgg Trp	acc Thr 340	Tyr	caa Gln	atc Ile	tac Tyr	cag Gln 345	gag Glu	ccc Pro	ttc Phe	aag Lys	aac Asn 350	1059
ctg Leu	aag Lys	act Thr	ggc Gly	aag Lys 355	tat Tyr	gcc Ala	agg Arg	atg Met	agg Arg 360	Gly ggg	gcc Ala	cac His	acc Thr	aat Asn 365	gat Asp	1107
gtg Val	aag Lys	cag Gln	ctg Leu 370	act Thr	gag Glu	gct Ala	gtg Val	cag Gln 375	aag Lys	atc Ile	acc Thr	act Thr	gag Glu 380	tcc Ser	att Ile	1155
gtg Val	atc Ile	tgg Trp 385	ggc Gly	aag Lys	acc Thr	ccc Pro	aag Lys 390	ttc Phe	aag Lys	ctg Leu	ccc Pro	atc Ile 395	cag Gln	aag Lys	·gag Glu	1203
acc Thr	tgg Trp 400	gag Glu	acc Thr	tgg Trp	tgg Trp	act Thr 405	gag Glu	tac Tyr	tgg Trp	cag Gln	gcc Ala 410	acc Thr	tgg Trp	atc Ile	cct Pro	1251
gag Glu 415	tgg Trp	gag Glu	ttt Phe	gtg Val	aac Asn 420	acc Thr	ccc Pro	ccc Pro	ctg Leu	gtg Val 425	aag Lys	ctg Leu	tgg Trp	tac Tyr	cag Gln 430	1299
ctg Leu	gag Glu	aag Lys	gag Glu	ccc Pro 435	att Ile	gtg Val	Gly	gct Ala	gag Glu 440	acc Thr	ttc Phe	tat Tyr	gtg Val	gct Ala 445	Gly ggg	1347
gct Ala	gcc Ala	aac Asn	agg Arg 450	gag Glu	acc Thr	aag Lys	ctg Leu	ggc Gly 455	aag Lys	gct Ala	ggc	tat Tyr	gtg Val 460	acc Thr	aac Asn	1395
agg Arg	ggc Gly	agg Arg 465	cag Gln	aag Lys	gtg Val	gtg Val	acc Thr 470	Leu	act Thr	gac Asp	acc Thr	acc Thr 475	aac Asn	cag Gln	aag Lys	1443

			cag Gln													1491
			gtg Val													1539
			cag Gln													1587
			aag Lys 530												aag Lys	1635
			ggc Gly													1683
			ctg Leu													1731
			cac His													1779
ccc			gtg Val													1827
			gag Glu 610												atc Ile	1875
tgg Trp	cag Gln	ctg Leu 625	gcc Ala	tgc Cys	acc Thr	cac His	ctg Leu 630	gag Glu	ggc Gly	aag Lys	gtg Val	atc Ile 635	ctg Leu	gtg Val	gct Ala	1923
gtg Val	cat His 640	gtg Val	gcc Ala	tcc Ser	ggc Gly	tac Tyr 645	att Ile	gag Glu	gct Ala	gag Glu	gtg Val 650	atc Ile	cct Pro	gct Ala	gag Glu	1971
aca Thr 655	ggc Gly	cag Gln	gag Glu	act Thr	gcc Ala 660	tac Tyr	ttc Phe	ctg Leu	ctg Leu	aag Lys 665	ctg Leu	gct Ala	ggc Gly	agg Arg	tgg Trp 670	2019
			acc Thr													2067
aca Thr	gtg Val	agg Arg	gct Ala 690	gcc Ala	tgc Cys	tgg Trp	tgg Trp	gct Ala 695	Gly ggc	atc Ile	aag Lys	cag Gln	gag Glu 700	ttt Phe	Gly	2115
atc Ile	ccc Pro	tac Tyr 705	aac Asn	ccc Pro	cag Gln	tcc Ser	cag Gln 710	ggg Gly	gtg Val	gtg Val	gcc Ala	tcc Ser 715	atg Met	aac Asn	aag Lys	2163

gag Glu	ctg Leu 720	aag Lys	aag Lys	atc Ile	att Ile	ggg Gly 725	cag Gln	gtg Val	agg Arg	gac Asp	cag Gln 730	gct Ala	gag Glu	cac His	ctg Leu	2211
aag Lys 735	aca Thr	gct Ala	gtg Val	cag Gln	atg Met 740	gct Ala	gtg Val	ttc Phe	atc Ile	cac His 745	aac Asn	ttc Phe	aag Lys	agg Arg	aag Lys 750	2259
ggg Gly	ggc Gly	atc Ile	Gly ggg	ggc Gly 755	tac Tyr	tcc Ser	gct Ala	Gly aaa	gag Glu 760	agg Arg	att Ile	gtg Val	gac Asp	atc Ile 765	att Ile	2307
gcc Ala	aca Thr	gac Asp	atc Ile 770	cag Gln	acc Thr	aag Lys	gag Glu	ctc Leu 775	cag Gln	aag Lys	cag Gln	atc Ile	acc Thr 780	aag Lys	atc Ile	2355
cag Gln	aac Asn	ttc Phe 785	agg Arg	gtg Val	tac Tyr	tac Tyr	agg Arg 790	gac Asp	tcc Ser	agg Arg	aac Asn	ccc Pro 795	ctg Leu	tgg Trp	aag Lys	2403
ggc Gly	cct Pro 800	gcc Ala	aag Lys	ctg Leu	ctg Leu	tgg Trp 805	aag Lys	GJÅ aaa	gag Glu	ggg	gct Ala 810	gtg Val	gtg Val	atc Ile	cag Gln	2451
gac Asp 815	aac Asn	tct Ser	gac Asp	atc Ile	aag Lys 820	gtg Val	gtg Val	ccc Pro	agg Arg	agg Arg 825	aag Lys	gcc Ala	aag Lys	atc Ile	atc Ile 830	2499
agg Arg	gac Asp	tat Tyr	ggc Gly	aag Lys 835	cag Gln	atg Met	gct Ala	Gly aaa	gat Asp 840	gac Asp	tgt Cys	gtg Val	gcc Ala	tcc Ser 845	agg Arg	2547
			gac Asp 850		agc	ccgg	gca	gatc	t							2577
<pre><21 <21 <21</pre>	Asp 0> 4 1> 8 2> P	Glu 50 RT	Asp 850	*				gatc .rus-								2577
<pre><21 <21 <21 <21 <40 Met</pre>	Asp 0> 4 1> 8 2> P 3> H 0> 4 Ala	Glu 50 RT uman Pro	Asp 850 Imm	* unod Ser	efic Pro	ienc Ile	y Vi Glu	rus- Thr	1 Val 10					ΤЭ	Pro	2577
<pre><21 <21 <21 <21 <40 Met</pre>	Asp 0> 4 1> 8 2> P 3> H 0> 4 Ala	Glu 50 RT uman Pro	Asp 850	* unod Ser	efic Pro	ienc Ile	y Vi Glu	rus- Thr	1 Val 10					ΤЭ	Pro Lys	2577
<pre><21 <21 <21 <21 <40 Met 1 Gly</pre>	Asp 0> 4 1> 8 2> P 3> H 0> 4 Ala	Glu 50 RT uman Pro Asp	Asp 850 Imm	unod Ser 5	efic Pro Lys	ienc Ile Val	y Vi Glu Lys Cys	rus- Thr Gln 25	1 Val 10 Trp	Pro	Leu	Thr Lys	Glu 30	Glu		2577
<pre><21 <21 <21 <21 <40 Met 1 Gly Ile</pre>	0> 4 1> 8 2> P 3> H 0> 4 Ala Met Lys	Glu 50 RT uman Pro Asp	Asp 850 Imm Ile Gly 20 Leu	unod Ser 5 Pro	efic Pro Lys Glu	ienc Ile Val Ile Glu	y Vi Glu Lys Cys	rus- Thr Gln 25 Thr	Val 10 Trp Glu	Pro Met	Leu Glu Thr	Thr Lys 45	Glu 30 Glu	Glu Gly	Lys	2577
<pre>Gln <21 <21 <21 <21 <10 Met 1 Gly Ile Ile</pre>	Asp 0> 4 1> 8 2> P 3> H 0> 4 Ala Met Lys Ser 50	Glu 50 RT uman Pro Asp Ala 35 Lys	Asp 850 Imm Ile Gly 20 Leu	unod Ser 5 Pro Val	efic Pro Lys Glu Pro	ienc Ile Val Ile Glu 55	y Vi Glu Lys : Cys : 40 : Asr	Tus- Thr Gln 25 Thr	Val 10 Trp Glu	Pro Met Asn Lys	Leu Glu Thr 60	Thr Lys 45	Glu 30 Glu Val	Glu Gly Phe	Lys Lys	2577
<pre>Gln <21 <21 <21 <40 Met 1 Gly Ile Ile </pre>	Asp 0> 4 1> 8 2> P 3> H 0> 4 Ala Met Lys Ser 50 Lys	Glu 50 RT uman Pro Asp Ala 35 Lys	Asp 850 Imm Ile Gly 20 Leu Ile	Ser 5 Pro Val	efic Pro Lys Glu Pro	ienc Ile Val Ile Glu 55 Thr	y Vi Glu Lys Cys 40 Asr	Tus- Thr Gln 25 Thr	Val 10 Trp Glu Tyr Arg	Pro Met Asn Lys 75	Leu Glu Thr 60 Leu	Thr Lys 45 Pro	Glu 30 Glu Val Asp	Glu Gly Phe	Lys Lys Ala Arg	2577
<pre><21 <21 <21 <21 <40 Met 1 Gly Ile Ile 65 Glu</pre>	Asp 0> 4 1> 8 2> P 3> H 0> 4 Ala Met Lys Ser 50 Lys	Fro Asp Ala 35 Lys Lys Asr	Asp 850 Imm Ile Gly 20 Leu Lys Lys	Ser 5 Pro Val Gly Asp 85 Arg 61	efic Pro Lys Glu Pro Ser 70	ienc Ile Val Ile Glu 55 Thr	y Vi Glu Lys 20 40 Asr Lys	Thr Gln 25 Thr Pro Trp Phe	Val 10 Trp Glu Tyr Arg	Pro Met Asn Lys 75 Glu	Leu Glu Thr 60 Leu Val	Lys 45 Pro Val	Glu 30 Glu Val Asp	Gly Gly Phe Phe Gly 95 Let	Lys Lys Ala Arg 80	25//
<pre></pre>	Asp 0> 4 1> 8 2> P 3> H 0> 4 Ala Met Lys Ser 50 Lys Leu His	Fro Asp Ala Ala Ala Lys Lys Asr Pro	Asp 850 Imm Ile Gly 20 Leu Lys Lys Ala 100 Ala	Ser 5 Pro Val Gly Asp 85 Gly	efic Pro Lys Glu Pro Ser 70 Thr	ienc Ile Val Ile Glu 55 Thr Glr	Glu Lys (Cys 40 (Asr Lys Lys	Thr Gln 25 Thr Pro Trp Phe	Val 10 Trp Glu Tyr Arg Trp 90 Lys	Pro Met Asn Lys 75 Glu	Leu Glu Thr 60 Leu Val	Lys 45 Pro Val Glr	Glu 30 Glu Val Asp Leu Val 110 Phe	Gly Gly Phe Gly Gly Gly Gly Gly	Lys Lys Ala Arg 80 Ile	2577

Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Ala Ala Leu Tyr Val Gly
180 185 190 Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly Leu Thr Thr Pro Asp Lys Lys His Gln 210 215 220 Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys 225 230 235 240 Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys 355 360 365 Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp 385 390 395 Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp
405
416 Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Ala Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Ala Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Ala Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Gln Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile
515
520
525 Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ala Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Asp Glu His Glu Lys
565 570 575 Tyr His Ser Asn Trp Arg Ala Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Ala Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val His

PCT/US00/34724

WO 01/45748

```
Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr Gly
                                                             655
                                       650
                 645
Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro Val
                                                         670
                                 665
            660
Lys Thr Ile His Thr Ala Asn Gly Ser Asn Phe Thr Gly Ala Thr Val
                                                    685
        675
                             680
Arg Ala Ala Cys Trp Trp Ala Gly Ile Lys Gln Glu Phe Gly Ile Pro
                         695
                                                700
Tyr Asn Pro Gln Ser Gln Gly Val Val Ala Ser Met Asn Lys Glu Leu
                                                                 720
                                           715
                      710
705
Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys Thr
                                                             735
                 725
                                       730
Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly Gly 740 745 750
Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Val Asp Ile Ile Ala Thr
                             760
        755
Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln Asn
                                                780
                        775
Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro
785
                      790
                                            795
                                                                  800
Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn
                                                            815
               805
                                       810
Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp
820 825 830
            820
                                  825
Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp
                               840
Glu Asp
    850
<210> 5
<211> 2650
<212> DNA
<213> Human Immunodeficiency Virus-1
<220>
<221> CDS
<222> (8)...(2635)
<400> 5
gatcacc atg gat gca atg aag aga ggg ctc tgc tgt gtg ctg ctg ctg
                                                                             49
        Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu 1 5 10
tgt gga gca gtc ttc gtt tcg ccc agc gag atc tcc gcc ccc atc tcc
Cys Gly Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ala Pro Ile Ser
                                                                             97
ccc att gag act gtg cct gtg aag ctg aag cct ggc atg gat ggc ccc
Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro
                                                                            145
                                                                            193
aag gtg aag cag tgg ccc ctg act gag gag aag atc aag gcc ctg gtg
Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val
                                    55
gaa atc tgc act gag atg gag aag gag ggc aaa atc tcc aag att ggc
                                                                            241
Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly
ccc gag aac ccc tac aac acc cct gtg ttt gcc atc aag aag aag gac
                                                                            289
Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp
                                                  90
      80
```

				agg Arg													337
				tgg Trp 115													385
ctg Leu	aag Lys	aag Lys	aag Lys 130	aag Lys	tct Ser	gtg Val	act Thr	gtg Val 135	ctg Leu	gat Asp	gtg Val	ggg Gly	gat Asp 140	gcc Ala	tac Tyr		433
ttc Phe	tct Ser	gtg Val 145	ccc Pro	ctg Leu	gat Asp	gag Glu	gac Asp 150	ttc Phe	agg Arg	aag Lys	tac Tyr	act Thr 155	gcc Ala	ttc Phe	acc Thr		481
				aac Asn													529
				ggc Gly													577
atg Met	acc Thr	aag Lys	atc Ile	ctg Leu 195	gag Glu	ccc Pro	ttc Phe	agg Arg	aag Lys 200	cag Gln	aac Asn	cct Pro	gac Asp	att Ile 205	gtg Val		625
				atg Met													673
				acc Thr													721
tgg Trp	ggc Gly 240	ctg Leu	acc Thr	acc Thr	cct Pro	gac Asp 245	aag Lys	aag Lys	cac His	cag Gln	aag Lys 250	gag Glu	ccc Pro	ccc Pro	ttc Phe		769
ctg Leu 255	tgg Trp	atg Met	ggc Gly	tat Tyr	gag Glu 260	ctg Leu	cac His	ccc Pro	gac Asp	aag Lys 265	tgg Trp	act Thr	gtg Val	cag Gln	ccc Pro 270		817
att Ile	gtg Val	ctg Leu	cct Pro	gag Glu 275	aag Lys	gac Asp	tcc Ser	tgg Trp	act Thr 280	gtg Val	aat Asn	gac Asp	atc Ile	cag Gln 285	aag Lys		865
ctg Leu	gtg Val	Gly ggc	aag Lys 290	ctg Leu	aac Asn	tgg Trp	gcc Ala	tcc Ser 295	caa Gln	atc Ile	tac Tyr	cct Pro	ggc Gly 300	atc Ile	aag Lys		913
gtg Val	agg Arg	cag Gln 305	ctg Leu	tgc Cys	aag Lys	ctg Leu	ctg Leu 310	agg Arg	ggc Gly	acc Thr	aag Lys	gcc Ala 315	ctg Leu	act Thr	gag Glu		961
gtg Val	atc Ile 320	ccc Pro	ctg Leu	act Thr	gag Glu	gag Glu 325	gct Ala	gag Glu	ctg Leu	gag Glu	ctg Leu 330	gct Ala	gag Glu	aac Asn	agg Arg	-	1009

gag Glu 335	atc Ile	ct'g Leu	aag Lys	gag Glu	cct Pro 340	gtg Val	cat His	Gly ggg	gtg Val	tac Tyr 345	tat Tyr	gac Asp	ccc Pro	tcc Ser	aag Lys 350	1057
gac Asp	ctg Leu	att Ile	gct Ala	gag Glu 355	atc Ile	cag Gln	aag Lys	cag Gln	ggc Gly 360	cag Gln	Gly ggc	cag Gln	tgg Trp	acc Thr 365	tac Tyr	1105
caa Gln	atc Ile	tac Tyr	cag Gln 370	gag Glu	ccc Pro	ttc Phe	aag Lys	aac Asn 375	ctg Leu	aag Lys	act Thr	ggc Gly	aag Lys 380	tat Tyr	gcc Ala	1153
agg Arg	atg Met	agg Arg 385	Gly ggg	gcc Ala	cac His	acc Thr	aat Asn 390	gat Asp	gtg Val	aag Lys	cag Gln	ctg Leu 395	act Thr	gag Glu	gct Ala	1201
gtg Val	cag Gln 400	aag Lys	atc Ile	acc Thr	act Thr	gag Glu 405	tcc Ser	att Ile	gtg Val	atc Ile	tgg Trp 410	Gly ggc	aag Lys	acc Thr	ccc Pro	1249
aag Lys 415	ttc Phe	aag Lys	ctg Leu	ccc Pro	atc Ile 420	cag Gln	aag Lys	gag Glu	acc Thr	tgg Trp 425	gag Glu	acc Thr	tgg Trp	tgg Trp	act Thr 430	1297
gag Glu	tac Tyr	tgg Trp	cag Gln	gcc Ala 435	acc Thr	tgg Trp	atc Ile	cct Pro	gag Glu 440	tgg Trp	gag Glu	ttt Phe	gtg Val	aac Asn 445	acc Thr	1345
ccc Pro	ccc Pro	ctg Leu	gtg Val 450	aag Lys	ctg Leu	tgg Trp	tac Tyr	cag Gln 455	ctg Leu	gag Glu	aag Lys	gag Glu	ccc Pro 460	att Ile	gtg Val	1393
Gly	Ala	Glu 465		Phe	Tyr	Val	Asp 470	Gly	Ala	Ala	Asn	Arg 475	GIU	Thr	гĀг	1441
Leu	Gly 480	Lys	gct Ala	Gly	Tyr	Val 485	Thr	Asn	Arg	GIA	Arg 490	Gin	ьуѕ	vaı	Vai	1489
Thr 495	Leu	Thr	gac Asp	Thr	Thr 500	Asn	Gln	Lys	Thr	G1u 505	Leu	Gin	Ala	IIe	510	1537
Leu	Ala	Leu	. Gln	Asp 515	Ser	Gly	Leu	Glu	Va⊥ 520	Asn	TTE	val	Tnr	525		1585
cag Gln	tat Tyr	gcc	ctg Leu 530	GLy	atc Ile	atc Ile	cag Gln	gcc Ala 535	GIN	cct Pro	gat Asp	cag Gln	tct Ser 540	GIU	tct Ser	1633
Glu	. Leu	Val 545	. Asn	. Gln	. Ile	Ile	550	Gln	. Leu	: Ile	: Lys	ьуs 555	GIU	. гу	gtg Val	1681
tac Tyr	ctg Leu 560	Ala	tgg Trp	gtg Val	cct Pro	gcc Ala 565	His	aaç Lys	g ggc Gly	att Ile	ggg Gly 570	CHA	aat Asn	gag Glu	cag Gln	1729

	gac Asp																1777
	att Ile																1825
	gct Ala																1873
	gtg Val																1921
	cag Gln 640																1969
	gag Glu															-	2017
	gag Glu																2065
	ctg Leu																2113
	aat Asn																2161
	gct Ala 720														tcc Ser		2209
	Gly ggg																2257
	gtg Val																2305
gtg Val	ttc Phe	atc Ile	cac His 770	aac Asn	ttc Phe	aag Lys	agg Arg	aag Lys 775	Gly aaa	ggc Gly	atc Ile	GJA aaa	ggc Gly 780	tac Tyr	tcc Ser		2353
	gly ggg																2401
	ctc Leu 800																2449

```
agg gac tcc agg aac ccc ctg tgg aag ggc cct gcc aag ctg ctg tgg
                                                                    2497
Arg Asp Ser Arg Asn Pro Leu Trp Lys Gly Pro Ala Lys Leu Leu Trp
                                        825
aag ggg gag ggg gct gtg gtg atc cag gac aac tet gac atc aag gtg
                                                                    2545
Lys Gly Glu Gly Ala Val Val Ile Gln Asp Asn Ser Asp Ile Lys Val
                                    840
gtg ccc agg agg aag gcc aag atc atc agg gac tat ggc aag cag atg
                                                                    2593
Val Pro Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met
            850
gct ggg gat gac tgt gtg gcc tcc agg cag gat gag gac taa
                                                                    2635
Ala Gly Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp
                                                                    2650
agcccgggca gatct
<210> 6
<211> 875
<212> PRT
<213> Human Immunodeficiency Virus-1
<400> 6
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly 1 15
Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ala Pro Ile Ser Pro Ile
                                25
            20
Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val
                           40
       35
Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val Glu Ile
                      55
                                           60
    50
Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu
                    70
                                        75
Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr 85 90 95
              85
Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln
                               105
            100
Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys
                                                125
                            120
Lys Lys Ser Val Thr Val Leu Asp Val Gly Asp Ala Tyr Phe Ser
                        135
Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro
                                        155
                    150
Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu
                                                        175
                                    170
               165
Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr
                                                   190
                                185
            180
Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Asp Ile Val Ile Tyr
195 200 205
Gln Tyr Met Asp Asp Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln
                        215
    210
His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Leu Arg Trp Gly
                                        235
                    230
Leu Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp
                245
Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Val
                                                   270
                                265
            260
Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val
                                                285
                            280
Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg
```

```
Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile
                   310
                                        315
Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile
                                    330
                325
Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu
           340
                              345
Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln Trp Thr Tyr Gln Ile
                            360
Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met
   370
                       375
Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln
385
                    390
                                       395
Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe
               405
                                   410
Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr Glu Tyr
           420
                               425
                                                   430
Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro
       435
                           440
                                               445
Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala
    450
                                         460
                      455
Glu Thr Phe Tyr Val Asp Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly
                   470
                                       475
Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln Lys Val Val Thr Leu
485
490
495
              485
                                   490
                                                       495
Thr Asp Thr Thr Asn Gln Lys Thr Glu Leu Gln Ala Ile Tyr Leu Ala
                                                    510
            500
                                505
Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Asp Ser Gln Tyr
                          520
                                                525
Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Gln Ser Glu Ser Glu Leu
                       535
                                           540
Val Asn Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu
                   550
                                       555
Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp
               565
                                  570
Lys Leu Val Ser Ala Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile
           580
                               585
                                                   590
Asp Lys Ala Gln Asp Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala
595 600
                          600
                                               605
Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val
                       615
                                           620
Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln
                 630
                                        635
Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Asp Cys Thr His Leu Glu
               645
                                   650
                                                       655
Gly Lys Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu
           660
                              665
Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu
                            680
                                                685
Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr Ile His Thr Asp Asn
   690
                       695
                                           700
Gly Ser Asn Phe Thr Gly Ala Thr Val Arg Ala Ala Cys Trp Trp Ala
                    710
                                        715
Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly 725 730 735
Val Val Glu Ser Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val
            740
                                745
                                                    750
Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe
                            760
       755
                                                765
Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly 770 775 780
Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu
```

Gln	ГЛЗ	Gln	Ile	Thr 805	Lys	Ile	Gln	Asn	Phe 810	Arg	Val	Tyr	Tyr	Arg 815	Asp	
Ser	Arg	Asn	Pro 820	Leu	Trp	Lys	Gly	Pro 825	Ala	Lys	Leu	Leu	Trp 830	Lys	Gly	
Glu	Gly	Ala 835		Val	Ile	Gln	Asp 840		Ser	Asp	Ile	Lys 845	Val	Val	Pro	٠
Arg	Arg 850		Ala	Lys	Ile	Ile 855		Asp	Tyr	Gly	Lys 860	Gln	Met	Ala	Gly	
Asp 865		Cys	Val	Ala	Ser 870		Gln	Asp	Glu	Asp 875						
<212	L> 26 2> DN	ΙA	Imm	ınode	efici	.ency	y Vir	rus-1	L							
)> L> CI 2> (8		(263	35)							•					
<400 gate)> 7 cacc	atg Met 1	gat Asp	gca Ala	atg Met	aag Lys 5	aga Arg	Gly aaa	ctc Leu	tgc Cys	tgt Cys 10	gtg V al	ctg Leu	ctg Leu	ctg Leu	. 49
tgt Cys 15	gga Gly	gca Ala	gtc Val	ttc Phe	gtt Val 20	tcg Ser	ccc Pro	agc Ser	gag Glu	atc Ile 25	tcc Ser	gcc Ala	ccc Pro	atc Ile	tcc Ser 30	97
ccc Pro	att Ile	gag Glu	act Thr	gtg Val 35	cct Pro	gtg Val	aag Lys	ctg Leu	aag Lys 40	cct Pro	ggc Gly	atg Met	gat Asp	ggc Gly 45	ccc Pro	145
aag Lys	gtg Val	aag Lys	cag Gln 50	tgg Trp	ccc Pro	ctg Leu	act Thr	gag Glu 55	gag Glu	aag Lys	atc Ile	aag Lys	gcc Ala 60	ctg Leu	gtg Val	193
gaa Glu	atc Ile	tgc Cys 65	act Thr	gag Glu	atg Met	gag Glu	aag Lys 70	gag Glu	ggc Gly	aaa Lys	atc Ile	tcc Ser 75	aag Lys	att Ile	ggc Gly	241
ccc Pro	gag Glu 80	aac Asn	ccc Pro	tac Tyr	aac Asn	acc Thr 85	cct Pro	gtg Val	ttt Phe	gcc Ala	atc Ile 90	aag Lys	aag Lys	aag Lys	gac Asp	289
tcc Ser 95	acc Thr	aag Lys	tgg Trp	agg Arg	aag Lys 100	ctg Leu	gtg Val	gac Asp	ttc Phe	agg Arg 105	gag Glu	ctg Leu	aac Asn	aag Lys	agg Arg 110	337
acc Thr	cag Gln	gac Asp	ttc Phe	tgg Trp 115	gag Glu	gtg Val	cag Gln	ctg Leu	ggc Gly 120	atc Ile	ccc Pro	cac His	ccc Pro	gct Ala 125	ggc Gly	385
ctg Leu	aag Lys	aag Lys	aag Lys 130	aag Lys	tct Ser	gtg Val	act Thr	gtg Val 135	ctg Leu	gct Ala	gtg Val	ggg ggg	gat Asp 140	gcc Ala	tac Tyr	433
ttc Phe	tct Ser	gtg Val 145	ccc Pro	ctg Leu	gat Asp	gag Glu	gac Asp 150	ttc Phe	agg Arg	aag Lys	tac Tyr	act Thr 155	gcc Ala	ttc Phe	acc Thr	481

					aat Asn											529
					tgg Trp 180											577
					gag Glu											625
					gct Ala											673
					aag Lys											721
					cct Pro											769
					gag Glu 260										ccc Pro 270	817
					aag Lys											865
					aac Asn											913
					aag Lys											961
gtg Val	atc Ile 320	ccc Pro	ctg Leu	act Thr	gag Glu	gag Glu 325	gct Ala	gag Glu	ctg Leu	gag Glu	ctg Leu 330	gct Ala	gag Glu	aac Asn	agg Arg	1009
gag Glu 335	atc Ile	ctg Leu	aag Lys	gag Glu	cct Pro 340	gtg Val	cat His	ggg Gly	gtg Val	tac Tyr 345	tat Tyr	gac Asp	ccc Pro	tcc Ser	aag Lys 350	1057
					atc Ile											1105
caa Gln	atc Ile	tac Tyr	cag Gln 370	gag Glu	ccc Pro	ttc Phe	aag Lys	aac Asn 375	ctg Leu	aag Lys	act Thr	Gly ggc	aag Lys 380	tat Tyr	gcc Ala	1153
					cac His											1201

gtg Val	cag Gln 400	aag Lys	atc Ile	acc Thr	act Thr	gag Glu 405	tcc Ser	att Ile	gtg Val	atc Ile	tgg Trp 410	ggc Gly	aag Lys	acc Thr	ccc Pro	124	9
aag Lys 415	ttc Phe	aag Lys	ctg Leu	ccc Pro	atc Ile 420	cag Gln	aag Lys	gag Glu	acc Thr	tgg Trp 425	gag Glu	acc Thr	tgg Trp	tgg Trp	act Thr 430	129	7
gag Glu	tac Tyr	tgg Trp	cag Gln	gcc Ala 435	acc Thr	tgg Trp	atc Ile	cct Pro	gag Glu 440	tgg Trp	gag Glu	ttt Phe	gtg Val	aac Asn 445	acc Thr	134	5
ccc Pro	ccc Pro	ctg Leu	gtg Val 450	aag Lys	ctg Leu	tgg Trp	tac Tyr	cag Gln 455	ctg Leu	gag Glu	aag Lys	gag Glu	ccc Pro 460	att Ile	gtg Val	139	3
Gly ggg	gct Ala	gag Glu 465	acc Thr	ttc Phe	tat Tyr	gtg Val	gct Ala 470	Gly aga	gct Ala	gcc Ala	aac Asn	agg Arg 475	gag Glu	acc Thr	aag Lys	144	1
ctg Leu	ggc Gly 480	aag Lys	gct Ala	ggc Gly	tat Tyr	gtg Val 485	acc Thr	aac Asn	agg Arg	ggc Gly	agg Arg 490	cag Gln	aag Lys	gtg Val	gtg Val	148	9
acc Thr 495	ctg Leu	act Thr	gac Asp	acc Thr	acc Thr 500	aac Asn	cag Gln	aag Lys	act Thr	gcc Ala 505	ctc Leu	cag Gln	gcc Ala	atc Ile	tac Tyr 510	153	7
ctg Leu	gcc Ala	ctc Leu	cag Gln	gac Asp 515	tct Ser	ggc Gly	ctg Leu	gag Glu	gtg Val 520	aac Asn	att Ile	gtg Val	act Thr	gcc Ala 525	tcc Ser	158	5
cag Gln	tat Tyr	gcc Ala	ctg Leu 530	Gly	atc Ile	atc Ile	cag Gln	gcc Ala 535	cag Gln	cct Pro	gat Asp	cag Gln	tct Ser 540	gag Glu	tct Ser	· 163	3
gag Glu	ctg Leu	gtg Val 545	aac Asn	cag Gln	atc Ile	att Ile	gag Glu 550	cag Gln	ctg Leu	atc Ile	aag Lys	aag Lys 555	gag Glu	aag Lys	gtg Val	168	31
tac Tyr	ctg Leu 560	gcc Ala	tgg Trp	gtg Val	cct Pro	gcc Ala 565	cac His	aag Lys	ggc Gly	att Ile	ggg Gly 570	ggc	aat Asn	gag Glu	cag Gln	172	29
gtg Val 575	gac Asp	aag Lys	ctg Leu	gtg Val	tct Ser 580	gct Ala	ggc	atc Ile	agg Arg	aag Lys 585	gtg Val	ctg Leu	ttc Phe	ctg Leu	gat Asp 590	177	17
ggc Gly	att Ile	gac Asp	aag Lys	gcc Ala 595	cag Gln	gat Asp	gag Glu	cat His	gag Glu 600	гÀг	tac Tyr	cac His	tcc Ser	aac Asn 605	JIP	182	25
agg Arg	gct Ala	atg Met	gcc Ala 610	Ser	gac Asp	ttc Phe	aac Asn	ctg Leu 615	Pro	cct Pro	gtg Val	gtg Val	gct Ala 620	гуs	gag Glu	187	73
att Ile	gtg Val	gcc Ala 625	Ser	tgt Cys	gac Asp	aag Lys	tgc Cys 630	Glr	ctg Leu	aag Lys	Gly	gag Glu 635	. дта	atg Met	cat His	192	21

ggg cag gtg g Gly Gln Val A 640	gac tgc tcc Asp Cys Ser	cct ggc at Pro Gly Il 645	c tgg cag e Trp Gln	ctg gcc tg Leu Ala Cy 650	c acc cac s Thr His	1969
ctg gag ggc a Leu Glu Gly I 655	aag gtg atc Lys Val Ile 660	ctg gtg gc Leu Val Al	t gtg cat a Val His 665	gtg gcc tc Val Ala Se	c ggc tac r Gly Tyr 670	2017
att gag gct g Ile Glu Ala G						²⁰⁶⁵
ttc ctg ctg a Phe Leu Leu I	aag ctg gct Lys Leu Ala 690	ggc agg tg Gly Arg Tr 69	p Pro Val	aag acc at Lys Thr Il 70	e His Thr	2113
gcc aat ggc t Ala Asn Gly S 705						2161
tgg gct ggc a Trp Ala Gly 1 720	atc aag cag Ile Lys Gln	gag ttt gg Glu Phe Gl 725	c atc ccc y Ile Pro	tac aac cc Tyr Asn Pr 730	c cag tcc o Gln Ser	2209
cag ggg gtg g Gln Gly Val V 735	gtg gcc tcc Val Ala Ser 740	atg aac aa Met Asn Ly	g gag ctg s Glu Leu 745	aag aag at Lys Lys Il	c att ggg e Ile Gly 750	2257
cag gtg agg g Gln Val Arg A						2305
gtg ttc atc o			rs Gly Gly		y Tyr Ser	2353
gct ggg gag a Ala Gly Glu A 785	agg att gtg Arg Ile Val	gac atc at Asp Ile Il 790	t gcc aca le Ala Thr	gac atc ca Asp Ile Gl 795	g acc aag n Thr Lys	2401
gag ctc cag a Glu Leu Gln I 800						2449
agg gac tcc a Arg Asp Ser A 815	agg aac ccc Arg Asn Pro 820	ctg tgg aa Leu Trp Ly	ag ggc cct vs Gly Pro 825	gcc aag ct Ala Lys Le	g ctg tgg u Leu Trp 830	2497
aag ggg gag g Lys Gly Glu (ggg gct gtg Gly Ala Val 835	gtg atc ca Val Ile Gl	ag gac aac ln Asp Asn 840	tct gac at Ser Asp Il	c aag gtg e Lys Val 845	2545
gtg ccc agg a Val Pro Arg A	agg aag gcc Arg Lys Ala 850	aag atc at Lys Ile I1 85	le Arg Asp	tat ggc aa Tyr Gly Ly 86	s Gln Met	2593
gct ggg gat g Ala Gly Asp A 865					a ·	2635
agcccgggca ga	atct					2650

PCT/US00/34724

```
<210> 8
<211> 875
<212> PRT
<213> Human Immunodeficiency Virus-1
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
                                   10
Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ala Pro Ile Ser Pro Ile
                               25
           20
Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val
       35
Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val Glu Ile
                                           60
                        55
   50
Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu 65 70 75 80
Asn Pro Tyr Asn Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr
                                   90
Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln
                               105
                                                    110
           100
Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys
115 120 125
Lys Lys Lys Ser Val Thr Val Leu Ala Val Gly Asp Ala Tyr Phe Ser
130 135 140
   130
Val Pro Leu Asp Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro
                                       155
                  150
Ser Ile Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu
                                    170
                165
Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr
180 185 190
                             185
           180
Lys Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Asp Ile Val Ile Tyr
                            200
                                               205
        195
Gln Tyr Met Ala Ala Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln 210 215 220
His Arg Thr Lys Ile Glu Glu Leu Arg Gln His Leu Arg Trp Gly
                                                             240
                                       235
                  230
Leu Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp
                                    250
                245
Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Val
                                                    270
                                265
           260
Leu Pro Glu Lys Asp Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val
                            280
Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Pro Gly Ile Lys Val Arg
                                            300
                        295
Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile
                                                             320
                   310
                                       315
Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile
                                    330
                325
Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu 340 345 350
Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln Trp Thr Tyr Gln Ile
                             360
        355
Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met
                                             380
                        375
Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln
                                      395
                     390
Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe
                                                       415
                                    410
                405
Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr Glu Tyr
                                                    430
                               425.
            420
 Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro
```

```
Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala
    450
                        455
                                            460
Glu Thr Phe Tyr Val Ala Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly
465
                                       475
                   470
Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln Lys Val Val Thr Leu
                                   490
               485
                                                        495
Thr Asp Thr Thr Asn Gln Lys Thr Ala Leu Gln Ala Ile Tyr Leu Ala
            500
                               505
                                                   510
Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Ala Ser Gln Tyr
515 520
       515
                          520
                                               525
Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Gln Ser Glu Ser Glu Leu
                       535
                                           540
Val Asn Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu
                  550
                                      555
Ala Trp Val Pro Ala His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp
                                   570
               565
                                                       575
Lys Leu Val Ser Ala Gly Ile Arg Lys Val Leu Phe Leu Asp Gly Ile
           580
                               585
Asp Lys Ala Gln Asp Glu His Glu Lys Tyr His Ser Asn Trp Arg Ala
        595
                            600
Met Ala Ser Asp Phe Asn Leu Pro Pro Val Val Ala Lys Glu Ile Val
   610
                       615
                                           620
Ala Ser Cys Asp Lys Cys Gln Leu Lys Gly Glu Ala Met His Gly Gln 625
                  630
                                      -635
Val Asp Cys Ser Pro Gly Ile Trp Gln Leu Ala Cys Thr His Leu Glu
               645
                                   650
                                                       655
Gly Lys Val Ile Leu Val Ala Val His Val Ala Ser Gly Tyr Ile Glu
           660
                              665
                                                   670
Ala Glu Val Ile Pro Ala Glu Thr Gly Gln Glu Thr Ala Tyr Phe Leu
       675
                           680
                                               685
Leu Lys Leu Ala Gly Arg Trp Pro Val Lys Thr Ile His Thr Ala Asn
                       695
                                          700
Gly Ser Asn Phe Thr Gly Ala Thr Val Arg Ala Ala Cys Trp Trp Ala
705
                   710
                                       715
                                                           720
Gly Ile Lys Gln Glu Phe Gly Ile Pro Tyr Asn Pro Gln Ser Gln Gly
               725
                                   730
                                                       735
Val Val Ala Ser Met Asn Lys Glu Leu Lys Lys Ile Ile Gly Gln Val
740 745 750
Arg Asp Gln Ala Glu His Leu Lys Thr Ala Val Gln Met Ala Val Phe
       755
                           760
Ile His Asn Phe Lys Arg Lys Gly Gly Ile Gly Gly Tyr Ser Ala Gly 770 780
Glu Arg Ile Val Asp Ile Ile Ala Thr Asp Ile Gln Thr Lys Glu Leu
                   790
                                       795
Gln Lys Gln Ile Thr Lys Ile Gln Asn Phe Arg Val Tyr Tyr Arg Asp
                                   810
              805
                                                       815
Ser Arg Asn Pro Leu Trp Lys Gly Pro Ala Lys Leu Leu Trp Lys Gly
           820
                               825
                                                  830
Glu Gly Ala Val Val Ile Gln Asp Asn Ser Asp Ile Lys Val Val Pro
       835
                           840
                                             845
Arg Arg Lys Ala Lys Ile Ile Arg Asp Tyr Gly Lys Gln Met Ala Gly
   850
                       855
Asp Asp Cys Val Ala Ser Arg Gln Asp Glu Asp
                    870
<210> 9
```

<400> 9

tegegegttt eggtgatgae ggtgaaaace tetgacacat geageteeeg gagaeggtea

<211> 4945

<212> DNA

<213> E. coli (V1Jns-tpa)

cagettgtet	gtaagcggat	accaaaaaca	gacaagcccg	tcagggcgcg	tcagcgggtg	120
ttaacaaata	teggggggg	cttaactato	cggcatcaga	gcagattgta	ctgagagtgc	180
accatatoco	atataaata	ccgcacagat	gcgtaaggag	aaaataccgc	atcagattgg	240
ctattaggg	ttacatacat	totatocata	tcataatatg	tacatttata	ttggctcatg	300
taanaantta	coccetatt	racattratt	attgactagt	tattaatagt	aatcaattac	360
	attactage	gatategate	gttccgcgtt	acataactta	cootaaatoo	420
ggggccacta	guidataged	catatatyya	gccccgcgcc	tcaataatca	catatattcc	480
cccgcctggc	tgaccgccca	acgaeccecg	cccattgacg	ccaacaacya	tagggtttt	540
catagtaacg	ccaataggga	ctttccattg	acgtcaatgg	grggagrace	tacygtaaac	600
tgcccacttg	gcagtacatc	aagtgtatca	tatgccaagt	acgececcia	Ligacyticaa	
tgacggtaaa	tggcccgcct	ggcattatgc	ccagtacatg	accttatggg	actitectae	660
ttggcagtac	atctacgtat	tagtcatcgc	tattaccatg	gtgatgcggt	tttggcagta	720
catcaatggg	cgtggatagc	ggtttgactc	acggggattt	ccaagtctcc	accccattga	780
cgtcaatggg	agtttgtttt	ggcaccaaaa	tcaacgggac	tttccaaaat	gtcgtaacaa	840
ctccacccca	ttgacgcaaa	tgggcggtag	gcgtgtacgg	tgggaggtct	atataagcag	900
agctcgttta	gtgaaccgtc	agatcgcctg	gagacgccat	ccacgctgtt	ttgacctcca	960 -
tagaagacac	caggaccgat	ccaqcctccq	cggccgggaa	cggtgcattg	gaacgcggat	1020
tececatace	aagagtgacg	taaqtaccqc	ctatagactc	tataggcaca	cccctttggc	1080
tettatgeat	actatactat	ttttaactta	gggcctatac	acccccgctt	ccttatgcta	1140
taggtgatgg	tatagettag	cctataggtg	tgggttattg	accattattq	accactcccc	1200
tattootoac	catactttcc	attactaatc	cataacatgg	ctctttacca	caactatctc	1260
tattggtgat	ataccaatac	tctatccttc	agagactgac	acqqactctq	tatttttaca	1320
gastagaata	ccatttatta	tttacaaatt	cacatataca	acaacaccat	ccccataca	1380
ggatggggtt	attanagata	acatagasta	tccacgcgaa	tetegggtae	atatteegga	1440
cycaytttt	tetagacta	gcgcgggacc	tccacatccg	accetacte	ccatacctcc	1500
catgggetet	Leteeggtag	cggcggagct	ctcatacccg	tageceeggee	acttaccac.	1560
ageggeteat	ggtegetegg	eageteetty	ctcctaacag	tagaggccag	atetatatat	. 1620
agcacaatgc	ccaccaccac	cagtgtgccg	cacaaggccg	tggtggtagg	teergegeee	1680
gaaaatgagc	gtggagattg	ggctcgcacg	gctgacgcag	atggaagact	caayycaycy	1740
gcagaagaag	atgcaggcag	ctgagttgtt	gtattctgat	aagagtcaga	ggtaacteec	
gttgcggtgc	tgttaacggt	ggagggcagt	gtagtctgag	cagtactcgt	rgergeegeg	1800
cgcgccacca	gacataatag	ctgacagact	aacagactgt	tcctttccat	gggtcttttc	1860
tgcagtcacc	gtccttagat	caccatggat	gcaatgaaga	gagggctctg	ctgtgtgctg	1920
ctgctgtgtg	gagcagtctt	cgtttcgccc	agcgagatct	gctgtgcctt	ctagttgcca	1980
gccatctgtt	gtttgcccct	ccccgtgcc	ttccttgacc	ctggaaggtg	ccactcccac	2040
tatcctttcc	taataaaatg	aggaaattgc	atcgcattgt	ctgagtaggt.	gtcattctat	2100
tctagagaat	aggatagga	aggacagcaa	gggggaggat	tgggaagaca	atagcaggca	2160
tactagagat	acaataaact	ctatggccgc	tgcggccagg	tgctgaagaa	ttgacccggt	2220
tectectaga	ccagaaagaa	gcaggcacat	ccccttctct	gtgacacacc	ctgtccacgc	2280
ccctaattct	tagttccagc	cccactcata	ggacactcat	agctcaggag	ggctccgcct	2340
tcaatcccac	cccctaaagt	acttggaggg	gtatatacat	ccctcatcag	cccaccaaac	2400
caaacctagc	ctccaagagt	gggaagaaat	taaagcaaga	taggctatta	agtgcagagg	2460
cadaccage	ccctccaaca	tataaaaaaa	taatgagaga	aatcatagaa	tttcttccac	2520
ttcatcactc	actorctoo	tacactcaat	cgttcggctg	caacaaacaa	tatcagetea	2580
atassaggg	attactage	tatccacaca	atcaggggat	aacacaaaaa	agaacatgtg	2640
cccaaaggcg	gcaacacggc	catcataga	taaaaaggcc	acattactaa	cotttttcca	2700
agcaaaaggc	cagcaaaaagg	ccaggaaceg	aaatcgacgc	tcaactcaca	ggtggggaaa	2760
Laggetteege	atatasagat	agcaccacaa	tcccctgga	ancteceten	tacactetee	2820
cccgacayga	ctataaayat	accaggeget	atacaacttt	ctcccttcag	raegeceteec	2880
tgttccgacc	etgeegetta	ceggatacet	gtccgccttt	tacctccccgg	gaagegegge	2940
gettteteat	ageteaeget	gtaggtatet	cagttcggtg	caggicgic	gttttaaget	3000
gggctgtgtg	cacgaacccc	cegttcagee	cgaccgctgc	geettateeg	gtaattattg	3060
tcttgagtcc	aacccggtaa	gacacgactt	atcgccactg	geageageea	ctggtaacag	
gattagcaga	gcgaggtatg	taggcggtgc	tacagagttc	ttgaagtggt	ggeetaaeta	3120
cggctacact	agaagaacag	tatttggtat	ctgcgctctg	ctgaagccag	ttaccttcgg	3180
aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	grggrrrrr	3240
tgtttgcaag	cagcagatta	cgcgcagaaa	aaaaggatct	caagaagatc	ctttgatctt	3300
ttctacgggg	tctgacgctc	agtggaacga	aaactcacgt	taagggattt	tggtcatgag	3360
attatcaaaa	aggatettea	cctagatcct	tttaaattaa	aaatgaagtt	ttaaatcaat	3420
ctaaagtata	tatgagtaaa	cttggtctga	cagttaccaa	tgcttaatca	gtgaggcacc	3480
tateteageg	atctotctat	ttcgttcatc	catagttgcc	tgactcgggg	ggggggggcg	3540
ctgaggtctg	cctcgtgaag	aaggtgttgc	tgactcatac	caggcctgaa	tcgccccatc	3600
atccadccad	aaagtgaggg	agccacggtt	gatgagagct	ttgttgtagg	tggaccagtt	3660
agtasttta	aacttttact	ttgccacgga	acggtctgcg	ttotcoogaa	gatgcqtqat	3720
ctdatccttc	aactcaccaa	aagttcgatt	tattcaacaa	agccaccatc	ccatcaaatc	3780
- cyastette	uucccugcaa	~~2		555	_	

```
agcgtaatgc tctgccagtg ttacaaccaa ttaaccaatt ctgattagaa aaactcatcg
                                                                                    3840
 agcatcaaat gaaactgcaa tttattcata tcaggattat caataccata tttttgaaaa
                                                                                    3900
 agccgtttct gtaatgaagg agaaaactca ccgaggcagt tccataggat ggcaagatcc
                                                                                    3960
 tggtatcggt ctgcgattcc gactcqtcca acatcaatac aacctattaa tttcccctcg
                                                                                    4020
, tcaaaaataa ggttatcaag tgagaaatca ccatgagtga cgactgaatc cggtgagaat
                                                                                    4080
 ggcaaaagct tatgcatttc tttccagact tgttcaacag gccagccatt acgctcgtca tcaaaatcac tcgcatcaac caaaccgtta ttcattcgtg attgcgcctg agcgagacga
                                                                                    4140
                                                                                    4200
 aatacgcgat cgctgttaaa aggacaatta caaacaggaa tcgaatgcaa ccggcgcagg
                                                                                    4260
 aacactgcca gcgcatcaac aatattttca cctgaatcag gatattcttc taatacctgg
                                                                                    4320
 aatgctgttt tcccggggat cgcagtggtg agtaaccatg catcatcagg agtacggata aaatgcttga tggtcggaag aggcataaat tccgtcagcc agtttagtct gaccatctca
                                                                                    4380
                                                                                    4440
 totgtaacat cattggcaac gotacotttg coatgtttca gaaacaacto tggcgcatcg
                                                                                    4500
 ggcttcccat acaatcgata gattgtcgca cctgattgcc cgacattatc gcgagcccat
                                                                                    4560
 ttatacccat ataaatcagc atccatgttg gaatttaatc gcggcctcga gcaagacgtt tcccgttgaa tatggctcat aacacccctt gtattactgt ttatgtaagc agacagtttt
                                                                                    4620
                                                                                    4680
 attgttcatg atgatatatt tttatcttgt gcaatgtaac atcagagatt ttgagacaca
                                                                                    4740
 acgtggcttt cccccccc ccattattga agcatttatc agggttattg tctcatgagc ggatacatat ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc
                                                                                    4800
                                                                                    4860
 cgaaaagtgc cacctgacgt ctaagaaacc attattatca tgacattaac ctataaaaaat
                                                                                    4920
 aggcgtatca cgaggccctt tcgtc
                                                                                    4945
 <210> 10
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> oligonucleotide
 <400> 10
 ctatataagc agagctcgtt tag
                                                                                      23
 <210> 11
 <211> 30
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> oligonucleotide
 <400> 11
 gtagcaaaga tctaaggacg gtgactgcag
                                                                                      30
 <210> 12
 <211> 39
<212> DNA
 <213> Artificial Sequence
 <220>
 <223> oligonucleotide
 <400> 12
 gtatgtgtct gaaaatgagc gtggagattg ggctcgcac
                                                                                      39
 <210> 13
 <211> 39
<212> DNA
 <213> Artificial Sequence
 <220>
 <223> oligonucleotide
```

```
<400> 13
                                                                                   39
gtgcgagccc aatctccacg ctcattttca gacacatac
<210> 14
<211> 4432
<212> DNA
<213> E. coli (V1J plasmid)
<400> 14
tegegegttt eggtgatgae ggtgaaaace tetgacacat geageteeeg gagaeggtea
                                                                                   60
cagettgtet gtaageggat geegggagea gacaageeeg teagegggtg
                                                                                  120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc
                                                                                  180
accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcagattgg
                                                                                  240
ctattggcca ttgcatacgt tgtatccata tcataatatg tacatttata ttggctcatg
                                                                                  300
tccaacatta ccgccatgtt gacattgatt attgactagt tattaatagt aatcaattac
                                                                                  360
ggggtcatta gttcatagcc catatatgga gttccgcgtt acataactta cggtaaatgg
                                                                                  420
cccgcctggc tgaccgcca acgacccccg cccattgacg tcaataatga cgtatgttcc catagtaacg ccaataggga ctttccattg acgtcaatgg gtggagtatt tacggtaaac
                                                                                  480
                                                                                  540
                                                                                  600
tgcccacttg gcagtacatc aagtgtatca tatgccaagt acgccccta ttgacgtcaa
tgacggtaaa tggcccgcct ggcattatgc ccagtacatg accttatggg actttcctac ttggcagtac atctacgtat tagtcatcgc tattaccatg gtgatgcggt tttggcagta
                                                                                  660
                                                                                  720
catcaatggg cgtggatagc ggtttgactc acggggattt ccaagtctcc accccattga
                                                                                  780
cgtcaatggg agtttgtttt ggcaccaaaa tcaacgggac tttccaaaat gtcgtaacaa
                                                                                  840
ctccgccca ttgacgcaaa tgggcggtag gcgtgtacgg tgggaggtct atataagcag agctcgttta gtgaaccgtc agatcgcctg gagacgccat ccacgctgtt ttgacctcca
                                                                                  900
                                                                                  960
                                                                                 1020
tagaagacac cgggaccgat ccagcctccg cggccgggaa cggtgcattg gaacgcggat
teccegtgee aagagtgacg taagtacege etatagagte tataggeeea ecceettgge
                                                                                 1080
ttcttatgca tgctatactg tttttggctt ggggtctata cacccccgct tcctcatgtt
                                                                                 1140
ataggtgatg gtatagctta gcctataggt gtgggttatt gaccattatt gaccactccc
                                                                                 1200
ctattggtga cgatactttc cattactaat ccataacatg gctctttgcc acaactctct ttattggcta tatgccaata cactgtcctt cagagactga cacggactct gtattttac
                                                                                 1260
                                                                                 1320
aggatggggt ctcatttatt atttacaaat tcacatatac aacaccaccg tccccagtgc
                                                                                 1380
ccgcagtttt tattaaacat aacgtgggat ctccacgcga atctcgggta cgtgttccgg
                                                                                 1440
acatgggete tteteeggta geggeggage ttetacatee gageeetget eccatgeete cagegactea tggtegeteg geageteett geteetaaca gtggaggeea gaettaggea
                                                                                 1500
                                                                                 1560
cagcacgatg cccaccacca ccagtgtgcc gcacaaggcc gtggcggtag ggtatgtgtc tgaaaatgag ctcggggagc gggcttgcac cgctgacgca tttggaagac ttaaggcagc
                                                                                 1620
                                                                                 1680
ggcagaagaa gatgcaggca gctgagttgt tgtgttctga taagagtcag aggtaactcc cgttgcggtg ctgttaacgg tggagggcag tgtagtctga gcagtactcg ttgctgccgc
                                                                                 1740
                                                                                 1800
gegegecace agacataata getgacagae taacagaetg tteettteea tgggtetttt
                                                                                 1860
                                                                                 1920
ctgcagtcac cgtccttaga tctgctgtgc cttctagttg ccagccatct gttgtttgcc
1980
                                                                                 2040
ggcagcacag caagggggag gattgggaag acaatagcag gcatgctggg gatgcggtgg gctctatggg tacccaggtg ctgaagaatt gacccggttc ctcctgggcc agaaagaagc
                                                                                 2100
                                                                                  2160
                                                                                 2220
aggcacatcc cettetetgt gacacaccet gtecacgece etggttetta gttecagece
cactcatagg acactcatag ctcaggaggg ctccgccttc aatcccaccc gctaaagtac
                                                                                 2280
ttggagcggt ctctcctcc ctcatcagcc caccaaacca aacctagcct ccaagagtgg
                                                                                 2340
gaagaaatta aagcaagata ggctattaag tgcagaggga gagaaaatgc ctccaacatg
                                                                                 2400
tgaggaagta atgagagaaa tcatagaatt tcttccgctt cctcgctcac tgactcgctg
                                                                                  2460
cgctcggtcg ttcggctgcg gcgagcggta tcagctcact caaaggcggt aatacggtta
                                                                                  2520
tccacagaat caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc
                                                                                  2580
                                                                                  2640
aggaaccgta aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgacgag
catcacaaaa atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac
                                                                                  2700
caggogtttc cocctggaag ctccctcgtg cgctctcctg ttccgaccct gccgcttacc
                                                                                  2760
ggatacetgt cegeetttet ecetteggga agegtggege tttetcaatg etcaegetgt
                                                                                  2820
aggtatetea gtteggtgta ggtegttege tecaagetgg getgtgtgea egaaceeee gtteageeeg acegetgege ettateeggt aactategte ttgagteeaa eeeggtaaga
                                                                                  2880
                                                                                  2940
cacgacttat cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta
                                                                                  3000
ggcggtgcta cagagttctt gaagtggtgg cctaactacg gctacactag aaggacagta
                                                                                  3060
                                                                                  3120
 tttggtatct gcgctctgct gaagccagtt accttcggaa aaagagttgg tagctcttga
 tccggcaaac aaaccaccgc tggtagcggt ggtttttttg tttgcaagca gcagattacg
                                                                                  3180
```

```
cgcagaaaaa aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag
                                                                                    3240
tggaacgaaa actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc
                                                                                    3300
tagatccttt taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact
                                                                                    3360
tggtctgaca gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggctta
                                                                                    3420
                                                                                    3480
ccatctggcc ccagtgctgc aatgataccg cgagacccac gctcaccggc tccagattta
                                                                                    3540
tcagcaataa accagccagc cggaagggcc gagcgcagaa gtggtcctgc aactttatcc
                                                                                    3600
gcctccatcc agtctattaa ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca acgttgttgc cattgctaca ggcatcgtgg tgtcacgctc gtcgtttggt
                                                                                    3660
                                                                                    3720
atggetteat teageteegg tteecaacga teaaggegag ttacatgate ecceatgitg
                                                                                    3780
tgcaaaaaag cggttagctc cttcggtcct ccgatcgttg tcagaagtaa gttggccgca
                                                                                    3840
gigttatcac teatggitat ggcagcactg cataaticte ttactgicat gccatccgta
                                                                                    3900
agatgetttt etgtgaetgg tgagtaetea accaagteat tetgagaata gtgtatgegg
                                                                                    3960
egacegagtt getettgeee ggegteaata egggataata eegegeeaca tageagaact
                                                                                    4020
ttaaaagtgc tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg
                                                                                    4080
ctgttgagat ccagttcgat gtaacccact cgtgcaccca actgatcttc agcatctttt
                                                                                    4140
actttcacca gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga
                                                                                    4200
ataagggcga cacggaaatg ttgaatactc atactcttcc tttttcaata ttattgaagc
                                                                                    4260
atttatcagg gttattgtct catgagegga tacatatttg aatgtattta gaaaaataaa caaatagggg tteegegeac attteecega aaagtgeeac etgaegteta agaaaceatt
                                                                                    4320
                                                                                    4380
attatcatga cattaaccta taaaaatagg cgtatcacga ggccctttcg tc
                                                                                    4432
<210> 15
<211> 4864
<212> DNA
<213> E. coli (V1Jneo plasmid)
<400> 15
tegegegttt eggtgatgae ggtgaaaace tetgacacat geageteeeg gagaeggtea
                                                                                      60
cagettgtet gtaageggat geegggagea gacaageeeg teagggegeg teagegggtg
                                                                                     120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc
                                                                                     180
accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcagattgg
                                                                                     240
ctattggcca ttgcatacgt tgtatccata tcataatatg tacatttata ttggctcatg tccaacatta ccgccatgtt gacattgatt attgactagt tattaatagt aatcaattac
                                                                                     300
                                                                                     360
ggggtcatta gttcatagcc catatatgga gttccgcgtt acataactta cggtaaatgg
                                                                                     420
cccgcctggc tgaccgccca acgacccccg cccattgacg tcaataatga cgtatgttcc
                                                                                     480
catagtaacg ccaataggga ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg gcagtacatc aagtgtatca tatgccaagt acgccccta ttgacgtcaa
                                                                                     540
                                                                                     600
tgacggtaaa tggcccgcct ggcattatgc ccagtacatg accttatggg actttcctac
                                                                                     660
ttggcagtac atctacgtat tagtcatcgc tattaccatg gtgatgcggt tttggcagta catcaatggg cgtggatagc ggtttgactc acggggattt ccaagtctcc accccattga cgtcaatggg agtttgttt ggcaccaaaa tcaacgggac tttccaaaat gtcgtaacaa
                                                                                     720
                                                                                     780
                                                                                     840
ctccgcccca ttgacgcaaa tgggcggtag gcgtgtacgg tgggaggtct atataagcag
                                                                                     900
agctcgttta gtgaaccgtc agatcgcctg gagacgccat ccacgctgtt ttgacctcca tagaagacac cgggaccgat ccagcctccg cggccgggaa cggtgcattg gaacgcggat
                                                                                     960
                                                                                    1020
teccegtgee aagagtgacg taagtacege ctatagagte tataggeeca eccettgge
                                                                                    1080
ttcttatgca tgctatactg tttttggctt ggggtctata cacccccgct tcctcatgtt
                                                                                    1140
ataggtgatg gtatagctta gcctataggt gtgggttatt gaccattatt gaccactccc ctattggtga cgatactttc cattactaat ccataacatg gctctttgcc acaactctct
                                                                                    1200
                                                                                    1260
ttattggcta tatgccaata cactgtcctt cagagactga cacggactct gtatttttac
                                                                                    1320
aggatggggt ctcatttatt atttacaaat tcacatatac aacaccaccg tccccagtgc
                                                                                    1380
ccgcagtttt tattaaacat aacgtgggat ctccacgcga atctcgggta cgtgttccgg
                                                                                    1440
acatgggete tteteeggta geggeggage ttetacatee gageeetget eccatgeete
                                                                                    1500
cagcgactca tggtcgctcg gcagctcctt gctcctaaca gtggaggcca gacttaggca
                                                                                    1560
cagcacgatg cccaccacca ccagtgtgcc gcacaaggcc gtggcggtag ggtatgtgtc tgaaaatgag ctcggggagc gggcttgcac cgctgacgca tttggaagac ttaaggcagc
                                                                                    1620
                                                                                    1680
ggcagaagaa gatgcaggca gctgagttgt tgtgttctga taagagtcag aggtaactcc
                                                                                    1740
cgttgcggtg ctgttaacgg tggagggcag tgtagtctga gcagtactcg ttgctgccgc gcgcgccacc agacataata gctgacagac taacagactg ttcctttcca tgggtctttt
                                                                                    1800
                                                                                    1860
ctgcagtcac cgtccttaga tctgctgtgc cttctagttg ccagccatct gttgtttgcc
                                                                                    1920
cetececegt geetteettg accetggaag gtgccactce caetgteett tectaataaa
                                                                                    1980
2040
```

PCT/US00/34724

```
ggcagcacag caagggggag gattgggaag acaatagcag gcatgctggg gatgcggtgg
                                                                                           2100
getetatggg tacceaggtg etgaagaatt gaceeggtte eteetgggee agaaagaage
                                                                                           2160
aggeacatec cettetett gacacacet gtecacgece etggttetta gttecagece cacteatagg acacteatag etcaggaggg etcegeette aateceacee getaaagtae
                                                                                           2220
                                                                                           2280
ttggagcggt ctctccctcc ctcatcagcc caccaaacca aacctagcct ccaagagtgg
                                                                                            2340
gaagaaatta aagcaagata ggctattaag tgcagaggga gagaaaatgc ctccaacatg tgaggaagta atgagagaaa tcatagaatt tcttccgctt cctcgctcac tgactcgctg
                                                                                           2400
                                                                                            2460
cgctcggtcg ttcggctgcg gcgagcggta tcagctcact caaaggcggt aatacggtta
                                                                                            2520
tccacagaat caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc
                                                                                            2580
aggaaccgta aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac
                                                                                            2640
                                                                                            2700
caggogtttc cccctggaag ctccctcgtg cgctctcctg ttccgaccct gccgcttacc
                                                                                            2760
ggatacetgt cegeetttet ecetteggga agegtggege ttteteaatg eteaegetgt
                                                                                            2820
aggtatctca gttcggtgta ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccc
                                                                                            2880
gttcagcccg accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga
                                                                                            2940
cacgacttat cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta
                                                                                            3000
ggcggtgcta cagagttctt gaagtggtgg cctaactacg gctacactag aaggacagta
tttggtatct gcgctctgct gaagccagtt accttcggaa aaagagttgg tagctcttga
tccggcaaac aaaccaccgc tggtagcggt ggtttttttg tttgcaagca gcagattacg
                                                                                            3060
                                                                                            3120
                                                                                            3180
cgcagaaaaa aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag
                                                                                            3240
tggaacgaaa actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact
                                                                                            3300
                                                                                            3360
tggtctgaca gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt
                                                                                            3420
cgttcatcca tagttgcctg actccggggg ggggggggcgc tgaggtctgc ctcgtgaaga aggtgttgct gactcatacc aggcctgaat cgccccatca tccagccaga aagtgaggga
                                                                                            3480
                                                                                            3540
gccacggttg atgagagett tgttgtaggt ggaccagttg gtgattttga acttttgett tgccacggaa cggtctgcgt tgtcgggaag atgcgtgate tgateettea actcagcaaa agttcgattt attcaacaaa gccgccgtce cgtcaagtca gcgtaatget ctgccagtgt
                                                                                            3600
                                                                                            3660
                                                                                            3720
tacaaccaat taaccaattc tgattagaaa aactcatcga gcatcaaatg aaactgcaat
                                                                                            3780
ttattcatat caggattatc aataccatat ttttgaaaaa gccgtttctg taatgaagga
                                                                                            3840
gaaaactcac cgaggcagtt ccataggatg gcaagatcct ggtatcggtc tgcgattccg actcgtccaa catcaataca acctattaat ttcccctcgt caaaaataag gttatcaagt
                                                                                            3900
                                                                                            3960
gagaaatcac catgagtgac gactgaatcc ggtgagaatg gcaaaagctt atgcatttct ttccagactt gttcaacagg ccagccatta cgctcgtcat caaaatcact cgcatcaacc
                                                                                            4020
                                                                                            4080
aaaccgttat tcattcgtga ttgcgcctga gcgagacgaa atacgcgatc gctgttaaaa
                                                                                            4140
ggacaattac aaacaggaat cgaatgcaac cggcgcagga acactgccag cgcatcaaca
                                                                                            4200
 atattttcac ctgaatcagg atattcttct aatacctgga atgctgtttt cccggggatc
                                                                                            4260
gcagtggtga gtaaccatgc atcatcagga gtacggataa aatgcttgat ggtcggaaga
                                                                                            4320
 ggcataaatt ccgtcagcca gtttagtctg accatctcat ctgtaacatc attggcaacg
                                                                                            4380
ctacctttgc catgtttcag aaacaactct ggcgcatcgg gcttcccata caatcgatag attgtcgcac ctgattgccc gacattatcg cgagcccatt tatacccata taaatcagca
                                                                                            4440
                                                                                            4500
 tocatgitgg aatttaatcg cggcctcgag caagacgttt cccgttgaat atggctcata
                                                                                            4560
 acacecettg tattactgtt tatgtaagca gacagtttta ttgttcatga tgatatattt
                                                                                            4620
                                                                                            4680
 ttatcttgtg caatgtaaca tcagagattt tgagacacaa cgtggctttc ccccccccc
 cattattgaa gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt
                                                                                            4740
 tagaaaaata aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc acctgacgtc
                                                                                            4800
 taagaaacca ttattatcat gacattaacc tataaaaata ggcgtatcac gaggcccttt
                                                                                            4860
                                                                                            4864
 cgtc
 <210> 16
 <211> 4867
 <212> DNA
 <213> E. coli (VlJns plasmid)
 <400> 16
 tegegegttt eggtgatgae ggtgaaaace tetgacacat geageteeeg gagaeggtea
                                                                                               60
 cagettgtet gtaageggat geegggagea gacaageeeg teagggegeg teagegggtg
                                                                                              120
 ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcagattgg
                                                                                              180
                                                                                              240
                                                                                              300
 ctattggcca ttgcatacgt tgtatccata tcataatatg tacatttata ttggctcatg
 tccaacatta ccgccatgtt gacattgatt attgactagt tattaatagt aatcaattac ggggtcatta gttcatagcc catatatgga gttccgcgtt acataactta cggtaaatgg
                                                                                              360
                                                                                              420
```

catagtaacg caataggga ctitccattg acgtcaatgg gtggagtait tacgciaaca (sqccccatt tgacgtcaac aggccccat ttgacgtcaac (aggcgtaat) gegtatatg ccagtacatg acgtcactat tggacgtaat (aggcggtaat) (aggcggatat) (aggcggatat) (aggcggatat) (aggggatt) (aggtggatat) (aggggatt) (aggtggatt) (aggtggatat) (agggggatt) (aggggatt) (agggggatt) (agggggatt) (agggggatt) (agggggatt) (agggggatt) (aggggggatt) (aggggggatt) (aggggggatt) (aggggggatt) (aggggggatt) (aggggggatt) (agggggggggggggggggggggggggggggggggggg							
catagtaacg caataggga ctitccattg acgtcaattg gtggagtatt tacggtaaa (gccccat ttgacgtcaa (agccgctaatta gactcaattg gagtacat aagtcagta (agcagtaattag) cagtaattag acttactag gactttcctac (accattga acttagga acttactga tagtcatcgc tattacattg gtgatgagggggggggg	cccacctaac	tgaccgccca	acqacccccq	cccattgacg	tcaataatga	cqtatqttcc	480
tgacgataa tgaccagca gaattataga catataga acgaccata ttgacgataa 600 tgaggataa tatacagtat tagtaataga catataga acgattacaa acttacagga atttacaata dettacagta tagtaataga cattataga cattataga cattacagta tagtaataga gattagataga ggattagata gagtagat tagtaagataga gattagataga							540
tgacggtaaa tggacggcat ggatatatgc cagtacatg accttatgg actttcctac foot ttggacgtaa acttacgtat tagtacatgc tattacatg gtgatgacggg ??? catcaatggg agtttpttt ggacacaaa cagggacttt ccaaatgt ctgataaagg agtttpttt ggacacaaa tagggaggtt gtggaggtct atgataagag 300 agctcgttt gtgacgcaaa tagggaggtag gggtgaaggaggaggaggaggaggaggaggaggaggagga							
tiggoagtaa atctacigtat tagtoatogo tattacoatg gfgatgoogt tittgocagta 720 catoaatgog ogtggatago ggttgatago ggttgatago ggttgatago ggttgatago ggttgatago ggttgatago ggttgatgog agttegatago ggttgatgog tittocaaaat gfogtaacaa 1900 agotogitta gfgaacqata agatogoctg ggaagagocat cacacgotgot titgacotca 960 tagaagaca caggatgag cagtocog oggocggat cocacctog oggocggat cocacctog caggatgago caggatgago ggttgatty gaacgogat 1020 tococgtgoc aagatgago tagtacogo catatagatca cacacctogot cottagagatg tagtagatgo gatatatagagot gagatgatagot catatagata accocctitiggo tatgaggatga gatacttac attacaata catacaatag cottitigoc attiggigagot attacatata catacaatag cottitigoc attiggigagot cocattatata attocaata totgoctot agagagotgot totataggata totacaata totgoctot agagagotgot cocattatata attocaata totgoctot agagagotgot cocattatata tattacaaat cacatacaa acaaccogaa totogggattit attaaacaata gogtgagat tocacaccaa acaacaccaa ggegogagot tocacaccaa agagocgot ggaagotgot cocacaccaa cagtgtgocg cacaagocog tagagotgo ggaagotgo gaaatago gtagagattg ggaccactag gtogagagot ggaagagotg ggaaatagag atgagagaatg ggaagaga tagagagaa gaaagaga atgaagaga cagagaga cagagagaga gaaagaga gtagagagag ggaagaga tagaagaga ggaaagaga gaaagaga agaagaga agaaga							
catcaatgog gyttgatagc gytttgactc acggggattt ccaagtctcc acccattga alloggaga agttigtttt gygaccaaa tggggaggaggaggaggaggaggaggaggaggaggaggag							
ggtcgtttt gtgaacgaa tggggggaa gcgtgatagg gtgggagta gggtgatagg gggagtga gggtgatagg gagtgatagg gagtgatagg gagtgatagg gagtgatagg gagtgatagg gagtgatagg gaacgcaa tataaggaag 900 agctogttt gtgaacgga caggtgatagg gagtgatagg gaacgcggaa coccettggaaggaa coccettggaaggaaggaaggaaggaaggaaggaaggaaggaag	ttggcagtac	atctacgtat	tagtcatcgc	tattaccatg	gtgatgcggt	tttggcagta	720
agctegtta grgaacegte agategecte gagacgecat cocacquette ttyaceteca 900 agctegtta grgaacegte coagcecteg eggecgggaa eggtqatatg gaacgegat teccegtgec aagagtgacg taagatacege ctatagacte tatagacac coccttagat tettatgeat getatactg ttitiggettg gggectatac accccettget 1080 tettatgeat getatactg ttitiggettg gggectatac accccettget cettatgeta tatggtatg tatagettag cetataggtg tgggatgattg accattattg accattactc tattggetat gatactttc attactaatc cataacatg ctetttgeca caactactc tattggetat gatactttc attactaatc cataacatga ctetttgeca caactactc ggatggggt caattatta ttacaaatt cacatacaca accacgegt cecegtgec ggatgggt cattatata ttacaaatt cacatacaca accacgegt cecegtgec ggatgggtc cattatata ttacaaatt cacatacaca accacgegt cocatgect cag ggatgggtc ttacacacac ggatggate tecacacegaa tetegggac cattaggac agcacaatg ccacacac cagtgtgec caacacacac ggaggace tacacacac ggaggacacacacacacacacacacacacac	catcaatggg	cgtggatagc	ggtttgactc	acggggattt	ccaagtctcc	accccattga	780
agctegtta grgaacegte agategecte gagacgecat cocacquette ttyaceteca 900 agctegtta grgaacegte coagcecteg eggecgggaa eggtqatatg gaacgegat teccegtgec aagagtgacg taagatacege ctatagacte tatagacac coccttagat tettatgeat getatactg ttitiggettg gggectatac accccettget 1080 tettatgeat getatactg ttitiggettg gggectatac accccettget cettatgeta tatggtatg tatagettag cetataggtg tgggatgattg accattattg accattactc tattggetat gatactttc attactaatc cataacatg ctetttgeca caactactc tattggetat gatactttc attactaatc cataacatga ctetttgeca caactactc ggatggggt caattatta ttacaaatt cacatacaca accacgegt cecegtgec ggatgggt cattatata ttacaaatt cacatacaca accacgegt cecegtgec ggatgggtc cattatata ttacaaatt cacatacaca accacgegt cocatgect cag ggatgggtc ttacacacac ggatggate tecacacegaa tetegggac cattaggac agcacaatg ccacacac cagtgtgec caacacacac ggaggace tacacacac ggaggacacacacacacacacacacacacac	catcaataaa	agtttgtttt	ggcaccaaaa	tcaacgggac	tttccaaaat	gtcgtaacaa	840
agcicgitta gigaacgic agaicgccig gagacgcat caacgigit tigaactca bagaagaaa coggaacquat caacacctoo gegacggaa coggaacquat caacacctoo gaacquat caacacctoo gaacquat caacacctoo caacaccac caacacctoo caacaccac caacaccaccac caacaccaccac caacaccac							900
tagaagaaca cgggaccgat caagactccg cggccgggaa cggtcattg gaacgcggat 1020 tectetageca aagataccg ctataagact tataagaca ccctttggc 1080 tettatgatg gctatactg ttttggcttg gggcctatac acccccgett ccttatgeta 1140 taggtgatgg tataacttac ttttggctat gggcctatac acccccgett ccttatgeta 1140 taggtgatgg tatagcttag ctataaggt ggggtattg acaattattg accaatcacc 1200 tattggtaa gatactttec attaaggta tgggtatgatg acaggactctg tattttaac 1260 tattggtaa atgccaatac tctgtcettc agagactgga caggactctg tattttaac 1260 tattggta decattatat tttacaaatt cacatataca acaaggcgt cccccgtgc 1130 agagggctct tetecagtaa gggtgggatc tecacatcga agccctggtc cattggcac 340 aggaccaatg ccacacaca cagttgtgcg cacaagaccag tgggggaag ttcacaatag 2 gaacaagac 2 gacacacac agttgtgcgg cacaagaccag tgggggaag ttaggtgcag agaaagaaga atggaagaag atggtgtgcag gggtgtgtgtgtggaaga tggtattgtgg gaagaagaag atggagaaga ggggagag tgatttgtg tatttgtga aagagtcag ggtagtgtg gggggaga tgattgtgg gaagaagaag atggagaaga ggggagag gggggag tgatttgga cactaccg tgctttggg gggggggggg							
tectagged aajagtgaeg taagtacege efatagaete tataggeae decettigee 1140 taggtgatgg tatagettag ectataggtg tgggetataa accecegete cettatgeta 1140 taggtgatgg tatagettag ectataggtg tgggetatat accecate cettetetete 1200 tattggtag gatactttee attactaate cataacatgg tetttgeae acactactec 1200 tattggtag gatactttee attactaate cataacatgg ectettigeae cacactactec 1200 gatgggggte coatttatta titacaaatt cacatataca acaacgeegt cecegityte 1240 catggggete tettetetete gaggaetgaat tettegggate tettetegate agggetett attacaaata tettegeggate tecacatgeag actettggggate tecacatgeag tetteggggate deagggetett etteteggg cagetettg tecacatgeag actetagggget gagaaggaagg gagaaggaagg egagaggate tecacatgeag accettggt catgggaaggaaggaaggagg gagaaggaaggagg gagaaggaaggagg							
tettageta gotafactgi tittiggetlig ggggetatae accecegett ecttageta 1140 taggigatgig tatagetlag estatagitg taggitatig acastatitg accactacce 1200 tattiggidat gatactitec attactaate cataacatig cittitigea caactatete 1260 tattiggeta datettee attactaate cataacatig cittitigea caactatete 1260 tattiggeta tatgecaata tetegetette agagactiga aggactety tattittaea 1320 ggatggggt coattatata titacaaaat cacatataca acaacacege eccecegige 1380 egaggetett tateacagta ggrigggate tecacateca agecetigge coatycete 1500 ageggeteat ggtegetegg eageteetty tecacateca agecetigge coatycete 1500 ageggeteat ggtegetegg eageteetty etcacateca gagecetigg eccapted 2560 ageggates eccacacae captiggeged agetageged gagetegeag gagagateggagateggagateggagateggagateggagateggagateggagagateggagagaga							
tatggtgatgg Tatagcttag cctataggtg töggttattg accattattg accattactc tattggtgac gatactttce attachaate cataacatgg ctctttgoca caacatactc 1260 tattggtat atgcaatac tctgtccttc agagactgac acggactctg tattttaca 1320 ggatggggtc ccattatatt attacaaate cacatactag acggactctg tattttaca 1320 catggggtctt attacacata tctgtccttc agagactgac acggactctg tattttaca 1320 catgggctctt attacacata ggatggggat tccacatacg accctggtc ccattgctg aggcggagt tccacatacg agcctggtc ccattgctgagggggggggg							
tattggtat atgocaata tetgetette agagactga cetettgoca caactatete 1260 gatggggte coattatata tetgetette agagactga caggactetg tattittaca 1320 gatgggggte coattatata tettacaaat cacatataca acaaggcgt ceccegigce 1380 catgggetet tetceoggtag gggtgggate tecacatega agecetggte catggetete 1500 agegggeteat ggtegetgg caggeggate tecacaataca gacactggte cataggeca 1500 ageggeteat ggtegetgg caggaggetgg catggegetgg caggagatetggag cacaggagetggggagaggaggaggaggaggaggaggagagaga	tcttatgcat	gctatactgt	ttttggcttg	gggcctatac	acccccgctt	ccttatgcta	1140
tattggtgat atgccaatac tctgtccttc agagactga cacgaactct batttttaca 1320 gattggggtc ccatttatta tttacaaatt cacatataca acaacgcgt cccccgtgcc gcagttttt attacacata gcgtgggatc tccaccgtac tccccgtggc cgcagttttt attacacata gcgtgggatc tccaccataca acaacgcgt cccccgtggc agcagactac ggtcgctcgg cagctcctty ctccacacac ggagaccga gtgttccgga agcacaatgc ccaccacac cagtgtgccg cacaaggccg tggaggcag acttaggcac gaaaatgagc gtgggagttg ggctcgcacg gaaaatgagc gtggagattg ggctcgcacg gaaaatgagc gtgtgaagttg ggctcgcacg gaaaatgagc gtgtacaggga ctggattgtt gtattctgat aagagtcaga gtaatccc tgcagtcacc gtctacaata gctgacagcag tgaagacgaa attgaagacat tacaccacg cacaaggcgt tccctccacaca gacacatatac tgcagtcacc gtccttagat ctgctgtgcc ttctagtgc cacaaggcgt tccttccat tggtctttc tgcagtcacc gtccttcatgat ctgctgtgcc ttctagtgc cacacaggaagaat gcagacgaaga gcagaaaaga atgcagagag ctgaagagac ttcaatcaca cactgtccttt cctaataacaa tgaaggacacc gtcctccttga ccctggaagag tgcagacacac gtcctccttga ccctggaagag gcaggacaaga aggggggggg attgggaaga gcaggacaagc atccccttct ctgtgacaca gcccacatca taggacacc atgccgcaca ggcccacacaca acacaccaca agcccacaca gacacacaca acaccacaca gacacacaca acaccacacacacacacacacacacacac	taggtgatgg	tatagettag	cctataggtg	tgggttattg	accattattg	accactcccc	1200
gattggggtc ccatttatta tittacaaatt cacatataca acaacgcctc coccetigcc 1380 cgcagttitt attaacacata gcgtgggatc tecacacgaa tetegggtac gtgttecga 1440 catgggetct teteacateca acaacgccg coccetigcc 1500 agcggetcta ggtgetcgg cagetcettg cecatecateca agcoctggtc ccategacac 1550 agcgcaatatgc cacaccac cagtgtgcc cacacacac tecategacac tecacacaca gagatgtgcc gacacacacacac cagtgtgccc cacacacacac gagatgtgcc cacacacacac gagatgtgcc cacacacacacacacacacacacacacacacacaca							1260
ggatggggtc ccatttatta ttracaaatt cacatataca acaacgccgt cccccgtgc cgagttttt attaaacata gcgtgggatc tccacacgcgaa tctcgggtac gtgttccgga 1440 catgggctct tctccggtag cggcggagct tccacatcg agccctggtc ccatggctcc 1500 agcgggtcat ggtggtcgg cagtccttg cccctaacag tggaggcaag acttaggcac agcacatgc ccacacac cagtgtgccg cacaaggcg tggaggcag acttaggcac gaaaatgag gtggagattg ggctcgcacg gctgacgcag atggaaggaa tatggtgtct fccgagcacaca gacataatag ctgacagagt gtagtctgag cagtacacga gtagaagaa tgaaggcag tgagagaaga gtagaggcag tgagagtag gtattctgat taggaagcag cgcgccacca gacataatag ctgacagact acaagactg tcctttccat gggtctttc cgcgccacca gacataatag ctgacagaag tgaagatcac tctttccat gggtcttttc ctccccgtg ccttccttga cctggaagg ttccatagac cacaagactg tcctttccat ggggtctttc ctccccgtg ccttccttga ccttgagtag tgcacatcc actgtccttt cctaataaaa 1980 gcaggacagc aagggggag attggaaga caatagcagg catgotgggg gtggggtggg gcaggaaat gcatcccttt cttgtaacac gcccactca taggacact atagctaag gacatcac acccttct cttgtaacac gccccactca taggacactc atagctcaag gtacttggg ggtgtctccc ctccctaat catgtgagg ggtactccc actgccaca accaaacca accaacca gcccactca taggacact atagctcaag gtacttgag ggttctccc ctccctaat aggacacg gtaggagag aataagaag gaatagaag agaatcaag gctgggagag agaaacaag gataggagag agaaacaag gcaggaagaa attaaagaa gataggagag agaaacaag gcaggaagaa attaaagaa gcaggaagaa attaagaag gcaggaagaa gcaggaagaa agaaacaag gcaggaagaa gcagaaaaag gcagaaaaag gcagaaaaag gcagaaaaag gcagaaaaaag gcagaaaaaaag gcagaaaaaaag gcagaaaaaaag gcagaaaaaaag gcagaaaaaaag gcagaaaaaaaa							
catgagetet beteeggtag eggegaget becaeggaa betegggtag eggtetee 1500 ageggeteat ggtegetegg eageteettg creatacag agecetggte ceatgegee 1500 ageaaatage ceaceaceae cagtgtgeeg cacaaageeg ggeggagag acttaaggeae 1560 ageaaatage gggagtag ggetegeeg cacaaageeg ggeggagag gatatgtgee 1680 geaaaatagae gtggagtatg ggetegeag getagaggaaatgagaeg gtaagetegae getagetegte gtatetetga aagageaag gtaageegg 1680 cgegecacea gacataatag etgacagaet gtatetetga aagagteaga ggtaatetee 1740 cgegecacea gacataatag etgacagaet gtatetetga aagagteaga ggtaatetee 1860 cgegecacea gacataatag etgacagaet aacagaetgt tecttecat gggtettte 1860 cgegecacea getettagat etgetggeeg tgecactee actgetett tectaataaaa 1980 cgaggaaat geategeatt gtetgagagg tgecactee actgetett tectaataaaa 1980 ctetatggee getgeggeea ggtgetgaag aattgagagg catgetgggg gtggggtggg 2040 ctetatggee getgeggeea ggtgetgaag aattgaceeg gttecteetg ggecagaaag cactgageaa attaageaa gataggeata actgetgggaga gatggegggag attggggagag attetgggagagaac actgeteetea cectgteeaa gaggeteege cettaateea ggeceacaca aaggegaga attaageaa gataggeata attaageag gatagetgag agggetegg 2100 cattggaga agtaatggaa gaaatcatag gatggeeagaa actaageaga gatageagaa agtaggagaa gataatgaag gaaatcatag gggagaagaa attaageaa gataggetat taagtgagaa gggaagaaa attaageaa gataggetat taagtgagaa gggaagaaa attaageaa gataggetat taagtgagaa gggaagaaa attaageaa gataggetat taagtgagaa gggaagaaa agtaagaga gataatgaag gataatgaag gggatagaa aggaagaaa gtaagagag agtatteeaa gaggagaa gataatgaga							
agregateat betreeggtag eggecettg cacetage ggecetggt coatgectec 1500 ageggeteat ggtegetegg cageteettg ctectaacag tggaggecag acttaggeac 1560 gaaaatgage gtggagaattg ggetegeag getageaga atggaagaet taaggeage 1680 ggeggagaagag atggaagagag atggagagag gtgagagag gtgagagaga							
aggactata ggtcgctcg caccacacac cagttgccg cacaaaggccg tgggggaag acttaggcac 1560 agacaatagc cacacacaca cagttgccg cacaaggccg tgggggaag gtatggtct 1620 gaaaatgaag gtagagaatt ggccgcacg gctgacgcag atggaagaat taaaggcaga gttgcggaag gtagtgcgc cacaagacca ggcagaagaag gttgcggcgcacaca gacataatag ctgagttgtt gtattctgat aagagtcag gtgagtcccc 1740 gcgccacac gacataatag ctgagtgtgt gtattctgat aagagtcag ggtactccc 1740 ctcccccgtg ccttccttga cctgagaga gtgcactccc actgtcata gggtcatttt cctccccgtg cctccccgtg cctccatga gtcggagag gtgcactccc actgtccatct gtgtttgcc tctaataaaa 1880 gcaggaagac aggaggaag gtcgagaga gtgcactccc actgtcact gtgttgcgc tctaatagac aggaggaaga gtcgagaga gtgcactccc actgtcacac ggccagaaga gtcgagaga gtcgagaga gtgcaactcc actgtgagag gtgcagagaga gtgcagagaga gtgcaactcc attgggaaga attggggag gtgtgaaga aattgaccag gtctgagag attgggagag attgggagag gtgcaacac actgtcacac ggccccagtaa gggcagaaa attaagcaa gagggcacac aggaggacac aggaggaga attagacac attgggaaga attaagaca gggcccacata agaacacac attagccaga gaaatcaaga gaggagaaa attaagcaaga gaaatcaaga gaaatcaaga gggaaacaca aaaaacaga cgtaaaaaaga gaaatcaag gggttatcaca gaatcagaga attaagaca gaaacacaa gacaacacaa acaaacaca gccccgaaaa gggagaaaa attaagcaaga gaaatcaaga gggttatcaca ggattaagac cactcaagaa gagaaacaaca gaaaacaaca gcaaaaaaga gaaacaaca ggagagaaa attaagcaaga gaaatcaaga gggttatcaca aaaaacaga ggaaatcaaga gaaacaacaa gaaaacaaca ggagagaaaa atcaagaga gaaatcaaga gaaacaacaa gaaaaacaaga ggaaacacaa aaaacaaca gcacaagaa aaaaacaga ggaaacacaa aaaacaacaa gaaaacaaca gaaaacaaca cacaaacaa							
agaaatgag gtgagattg ggctcgcag gctgacgag atgagagat taggagaagat tagaggaaga ctgagttgtt gtattctgat aagagtcaga ggtaactcc 1740 gttgcggtgc tgttaacggt gtgagttgtt gtattctgat aagagtcaga ggtaactcc 1740 gttgcggtgc tgttaacggt gtgagttgtt gtattctgat aagagtcaga ggtaactcc 1740 gttgcggtgc tgttaacggt gtgagtgtgt gtagtctgaa cagtactcgt tgtgcccccg 1800 ctcccccgtg ccttccttga cctgacagact aacagactgt tccttccat gggtcttttc 1860 tgaggaaatt gcatcgcatt gtctgagtag tgccatccc actgtccttt cctaataaa 1980 tgaggaaatt gcatcgcatt gtctgagtag gtgccatccc actgtccttt cctaataaaa 1980 ctcatcggacaga aagagggaga gtgctgagag gtgctgaaga attgaggaaga caatagcagg catgctgggg ggcgcaaaagagagagagagagagagagagagaga	catgggctct	tctccggtag	cggcggagct	tccacatccg	agccctggtc	ccatgcctcc	
gaaaatgag gtggagattg ggctcgag gctgacgag atggaagact baaggcaagg flass gaagaagaag atgcaggcag ctgagttgtt gtattctgat aagagtcaga ggtaactccc floor gttgcggtc byttaacgg ggagggcagt gtattctgat aagagtcaga ggtaactccc gcgcacaca gacataatag ctgacagact aacagactgt tcctttccat gggtcttttc tgagtcacc gcctctaaga tcgcgtcgct tctattgtc caccagactgt tccttcctg gggcctcccccccgtg ccttcctt							1560
gaaaatgag gtggagattg ggctcgag gctgacgag atggaagact baaggcaagg flass gaagaagaag atgcaggcag ctgagttgtt gtattctgat aagagtcaga ggtaactccc floor gttgcggtc byttaacgg ggagggcagt gtattctgat aagagtcaga ggtaactccc gcgcacaca gacataatag ctgacagact aacagactgt tcctttccat gggtcttttc tgagtcacc gcctctaaga tcgcgtcgct tctattgtc caccagactgt tccttcctg gggcctcccccccgtg ccttcctt	agcacaatgc	ccaccaccac	cagtgtgccg	cacaaggccg	tggcggtagg	gtatgtgtct	1620
gcagaagaag atgcaggcag ctgagttgtt gtattctgat aagagtcaga ggtaactccc gttgcggcgc tgttaacggt ggagggcagt gtagtctgag cagtactcgt tgctgccgcg cggccacca gacataatag ctgacagact aacagactgt tcctttcat gggtctttc 1860 ctcccccgtg ccttccttga ccctggaagg tgcactccc actgcctct cctaataaaa 1980 gcaggacagc aaggggagg attgggaaga caatagcagg catgcgggggggggg							1680
gttgoggtgc tgttaacggt ggagggcagt gtagtctgag cagtactcgt tgctgccgcg 1800 cgcgccacca gacataatag ctgacagact aacagactt tcctttccat gggtcttttc tgcagtcac gacataatag ctgcaggacgt tcgttagtt tcctttccat gggtcttttc tgcagtcac gtctcaccagt ctcttcattga ccctggaagg tgccatccc actgtccttt cctaataaaa 1980 tgaggacaac aagggggagg attgggaag caatagcagg catgctgggg gtgggggggggg							
cgcgcacaca gacataatag ctgacagact aacagactgt tcctttcat gggtttttc tgtgacacc gtccttagat ctgctgtgcc ttctagttgc caccactcgt ttgtttgcc ctcccccgtg ccttcattga ccctggaagg tgccactcc actgtccttt cctaataaaa 1980 gcaggaaatt gcatcgcatt gtctgaqtag tgctattct attctggggg gtggggggggaagaatt gcatcgcagac ggtgctgaag attgggaaga caatagcagg catggggg atgcggggggg cccactca taggacactc ctgtgacaca ccctgtccttc ggacagacag ggtgctgaag aattggaaga caatagcagg gtcccccctcat aggacactc ataggctcag ggtgcccactca aggacactc ctgtgacaca ccctgtccac gcccactcat aggacactc ataggctcag gagagcacg cttcaatcac acccgctaaa 2280 gtacttgggg gtaatagagg ataaggctat taagtgcaag aggagagaaa accacagctaa 2280 gtggggagga gtactggg gtgtggggggggggggggg							
tgcagtcacc gtccttagat ctgctgtgcc ttctagttgc cagccatctg ttgttgccc 1920 ctcccccgtg gtcctcttga ccctggaagg tgccactccc actgtcctt ctaataaaa 1980 gtaggaaatt gcatcgatt gtctagatag gtgccatcacc actgtcctt ctaataaaa 1980 ctctatggcc gaaggagga attgggaaga gtgccatcacc actgtctggg gtggggtggg							
ciccocgtg cettectiga cectggaagg tgccaetee actgteett cetaataaaa 1980 tyaggaaatt gcateggatgg gtgteatteet attetggggg gtgggtggg 2040 cetatggee getgeggeea gtgteaaa gtgteattet attetggggg gtgggtggg 2100 cetatggee getgeggeea gtgtegaag aattgggaaga caatageagg attgggaaga 2160 aagaagageae ateceettet cetgtgaacae cectgteeae geecetggtt ettagtteea geeceaetee atageteaag agggeteeae gggeeteagg gtacttegg gtacttggag ggaagaaaa atgaageag ataggetaat aaggeteaga gggaagaaa acaceggtaaa 2280 gtagtggga ggaatataga gaataggetat taagtgeaa gggaagaaaa atgacteeaa gaataggetat taagtgeag gggaagaaaa atgacteea gaataggetat taagtgeag gggaagaaaa atgacteea gaataggetaat aaattettee getteeteeg teactgaace 2400 getgegeeteg gtegteege tgeggegag ggtatteete cacteaaagg gggaagaaaa atgacagag ataacgaagg aaagaaacatg tgageaaaag geggaataace ggetaataeag aattetteeg ggggegage ggtattaceae gaataaaggag ceggttygt ggggttitte cataggetea geceectgg etteeege tteeteecetg ggaageteeet ggggggaa aacecgaaga accegacaaa 2580 gaagaacatea tegteegeet teeteecete gggaageteet cetteteege eectgeeget teeteegeet teeteegeet teeteegeet teeteegeet teeteegeet teeteegeet teeteegeet teeteegeet eegaagee gggaagagaa acacegaaga 2700 ataacaggata tetaggagat eegagteete gggaagegtgg gegettetee ataggeteae 2820 eegaagaaa agaacacag eegagteete gggaagegtgg gegetteete ataggeteae 2820 eegaagaa agaagaace eegaagaga 2700 ataacaggat tetegaagee gggaagagatag eegagetage 2820 eegaagaa 2700 ataacaggat tetegaagee gggaagagatag 2700 aagaagaagat eegagetag 2700 aagaagaaga 2700 aagaagaaga 2700 aagaagaaga 2700 aagaagaaga 2700 aagaagaagaa							
tgaggaaatt gcatcgcatt gtctgatag gtgtcattct attctggggg gtggggtggg							
gcaggacaga aagggggagg attggaaga caatagcagg catgctgggg attggagag 2100 aagcaggcac atcccttct ctgtgacaca catgctcgg gttcctcctg ggccagaaag 2160 gaccactca taggacactc ctgtgacaca ccctgtccac gcccctggtt cttagtcca gccccatca taggacactc atagctcagg agggctccgc cttcaatcc acccgctaaa 2280 gtacttggag cggtctctcc ctccctcatc agcccacaa accaaaccta gcctccaaga ctgtggagaaga attaaagcaa gataggctat taagtgcagg gggagagaaa atgcctccaa 2400 gctggcgctcg gtcgttcggc tgcggcaggc ggatacagc cactcaagac gggtaatacg gttatccaca gaatcaggg aaatcatag aatttcttcc gcttcctcgc tcactgactc 2460 gctggcgctcg gtcgttcggc tgcggcagg ggatacaggt cactcaaaga gggcacacaaa atgaacacgg ggtatacaag ggcaggaac cgtaaaaagg ccgcgttgct ggcggttgct ggcggttttc cactcaagag ggcacacaaa ggccacacaaa 2580 ggccaggaac cgtaaaaagg ccgcgttgct ggcggttttc cataggctcc gcccccctga 2640 cgagcatcac aaaaatcgac gctcaagtca gaggtgggga aacccgacag gactataaag 2700 ataccaggga tttccccctg gaagctccct cgtgcggctc cctgttccga ccctgccgct taccggtaac ctgtcagccct ttcccctcc gggaagcgg ggaagcggg gcgctttctc atagctcacg 2820 ctgtaggtat ctcagttcgg tgtaggtcgt tcgctccaag ctgggcttgt ggacgacc 2820 ctgtaggtat ctcagttcgg tgtaggtcgt tcgctccaag cgggtatacac 2820 ctgtaggcag ccagaccgc tggccttatc gggaactac cggcttggt ggaactac gggcttagg cactgggagac 2820 ctgtaggcggt gctacacagagt tcttgaagtg gtggcctaac aggattagga gagggaggga 3000 agtatttggt atctggccac tggcgtag cactggtaac aggattagac agagcagagt 120 aggacacga aaaaaacaca ccgctggtag gtggctaac tacggcaaa aagaacaca 2820 ctggtaggaa gaaaacaca ccgctggtag gtggctaac tacggcaaa 2820 ctggtggaa gaaaaaacaca ccgctggtag gtggcgaacac 2820 ctgtgtcgga aaaaaaacac ccgctggtag gtgggagga aaaccaggat 120 aggagaac ggacggcg gaaaacaca 2820 ctgtgtgaa gaaacacac 2820 ctgtgtgga aaacacacac 2820 cactggtaa cacacacacac 2820 cactggtaa cacacacacac 2820 cactggtaa cacacacacac 2820 cactggtaa cacacacacacacacacacacacacacacacacaca							1980
gcaggacaga aagggggagg attggaaga caatagcagg catgctgggg attggagag 2100 aagcaggcac atcccttct ctgtgacaca catgctcgg gttcctcctg ggccagaaag 2160 gaccactca taggacactc ctgtgacaca ccctgtccac gcccctggtt cttagtcca gccccatca taggacactc atagctcagg agggctccgc cttcaatcc acccgctaaa 2280 gtacttggag cggtctctcc ctccctcatc agcccacaa accaaaccta gcctccaaga ctgtggagaaga attaaagcaa gataggctat taagtgcagg gggagagaaa atgcctccaa 2400 gctggcgctcg gtcgttcggc tgcggcaggc ggatacagc cactcaagac gggtaatacg gttatccaca gaatcaggg aaatcatag aatttcttcc gcttcctcgc tcactgactc 2460 gctggcgctcg gtcgttcggc tgcggcagg ggatacaggt cactcaaaga gggcacacaaa atgaacacgg ggtatacaag ggcaggaac cgtaaaaagg ccgcgttgct ggcggttgct ggcggttttc cactcaagag ggcacacaaa ggccacacaaa 2580 ggccaggaac cgtaaaaagg ccgcgttgct ggcggttttc cataggctcc gcccccctga 2640 cgagcatcac aaaaatcgac gctcaagtca gaggtgggga aacccgacag gactataaag 2700 ataccaggga tttccccctg gaagctccct cgtgcggctc cctgttccga ccctgccgct taccggtaac ctgtcagccct ttcccctcc gggaagcgg ggaagcggg gcgctttctc atagctcacg 2820 ctgtaggtat ctcagttcgg tgtaggtcgt tcgctccaag ctgggcttgt ggacgacc 2820 ctgtaggtat ctcagttcgg tgtaggtcgt tcgctccaag cgggtatacac 2820 ctgtaggcag ccagaccgc tggccttatc gggaactac cggcttggt ggaactac gggcttagg cactgggagac 2820 ctgtaggcggt gctacacagagt tcttgaagtg gtggcctaac aggattagga gagggaggga 3000 agtatttggt atctggccac tggcgtag cactggtaac aggattagac agagcagagt 120 aggacacga aaaaaacaca ccgctggtag gtggctaac tacggcaaa aagaacaca 2820 ctggtaggaa gaaaacaca ccgctggtag gtggctaac tacggcaaa 2820 ctggtggaa gaaaaaacaca ccgctggtag gtggcgaacac 2820 ctgtgtcgga aaaaaaacac ccgctggtag gtgggagga aaaccaggat 120 aggagaac ggacggcg gaaaacaca 2820 ctgtgtgaa gaaacacac 2820 ctgtgtgga aaacacacac 2820 cactggtaa cacacacacac 2820 cactggtaa cacacacacac 2820 cactggtaa cacacacacac 2820 cactggtaa cacacacacacacacacacacacacacacacacaca	tgaggaaatt	gcatcgcatt	gtctgagtag	gtgtcattct	attctggggg	gtggggtggg	2040
ctctatggcc gctgcggca ggtgctgaag aattgacccg gttctctctg ggccagaaag 2160 aagcaggcaa atcccttct ctgtgacaca cccttctcac gcccctggtt cttagttcca gccccactca taggacactc ataggctcaga agggctcgcc cttcaatcca accgctaaa 2280 gtacttggag cggtctctcc ctccctcatc agcccacaa accaaaccta gcctccaag gtgggaagaa attaaagcaa gataggctat taagtgcaga gggaggagaaa atgcctccaa gctgggaaggaa gtaatgaga gaaatcatag aatttcttcc gcttcctcgc tcactgactc gttatccaca gatcagggg ataacggag ggatacagct cactcaaagg ggcaggaac cgtaaaaagg ccgcgttgct ggcgtttttc cataggctc dagagaacacggag aaagaacatg tgagcaaaag gccagcaaaa ggccaggaac cgtaaaaagg ccgcgttgct ggcgtttttc cataggcga accaggagac cgtcaagtca gagttgcga aacccgacag gactataaag ggccaggaac cttccccctg gaagctccct cgtgcgctc cctgtccgc cccctga ctgtaggtat ctcaggtcg ttccccctg gaggtgggga aacccgacag gactataaag cccggttagg tttccccctg gaagctcct cgtgcgctc cctgttccga ccctgccgct taccggatac ctgtccgcct ttccccctc gggaagcggg gcgctttctc atagctcacg ctgtaggtat ctcagttcgg tgtaggtcgt tcgctcaaag cggctttgt gcacgaaca cccgttcag cccgaccgct ggcgcttatc cggtaactat cggtctggg cacaccggt aagaacagaac ttatcgcac tggcagcag cactggtaac aggattagaa gaggaggtg aagaacaggac ttgcgcac tggcagacg cactggtaac aggattagaa gagagaggtg ttggtgaagaac caccgggtgggga aaccaggaggtg ttggtgaagaac ggtaacaagg tctgaaggag ggggggggggggggggggggggggggggg							2100
aagcaggaca atccectet cyggacaca cectytecae geeestgyt ettagteea 2220 geeecateaa taggacaete atageteagg agggeteege etteaateea accegetaaa 2280 gyaggaagaa attaaageaa gatagyatat taagtgeaga gyagagaaaa atgeeteeaa 2440 gtgggaagaa attaaageaa gatagyatat taagtgeaga gyagagaaaa atgeeteeaa 2440 getggegeteg gtegteegee tyeggegage gyatacagee cacteaaaagg cygtaataea 2520 gttateeaaa gaateagggg ataaacgaag aaagaacatg tyagagaaaaa geeagaaaaa 2580 gyeagagaaa egataaaagg eegetytget gyegytytee gyagttette catagyetee geeeecetya 2640 egaagaataea aaaaategae geeagatee gyagttytee gyagytytee gyagyagaaaa gyeaaaaa 2580 gyeagagaaa eegataaaagg eegetytget gyagytytee gyagytytee gyagyagaaaaa gyeaaaaag geeagagaa 2580 gyagagaaaa 2580 gyagagaaaaa aaaaaaagagagagagagagggggagagagagggggagag							2160
geoccactea taggacacte atageteagg agggeteege etteaateee accegetaaa 2280 gtacttygag eggeteteee eteceteate ageceaceaa accaaaceta geeteeaag 2340 catygaggaa aataaaggaa gaaaggetat taaagteaga gggagagaaa atgeeteeaa 2460 getgegeteg gtegtteege tgeeggagg ggatateage eacteeaagg eggeagagaee egetateee 2460 getaceacag agtacacgag gaaacacgg eggeagaeaa 2580 ggeeaggaac egtaaaaagg eeggetyget ggegttette eacteaaagg eggeageacacgag ataacgagg aaagaacatg tgagcaaaag geeageaaa 2580 ggeeaggaac egtaaaaagg eeggetyget ggegttette eactgaeaag geeageaaa 2580 eaceggetget tteeecetg gaageteeet egggaggegg aacceggaeg egageteeet eetgeggetee etgetgegete teetgaggetg egageteet eetgeggetee etgetgeget teeteeetg tgaggaggagg accegaaagg gactataaag 2700 gagaacacga ettateeetg tgtaggtegt egggetetet eetgegetee etgetgagget eeggeteet teeteetee gggaageegg geetteete etgetgeeget teeteetee eggagegggggggggg							
gtacttggag cggtctctcc ctccctcatc agccaccaa accaaaccta gcctccaaga gtgggaagaa attaaagcaa gataggctat taagtgcaga gggagagaaa atgcctccaa 2400 catgtgagga agtaatgaga gaaatcatag aattcttcc gcttcctcgc tcactgactc 2460 gctgcgctcg gtcgttcgc gtcgtcggc gggagagaaac aggcagaaca 2520 gttatccaca gaatcagggg ataacgagg ggatatcagct cactcaaagg cggtaatcac ggccaggaac cggtaaacagg ccgcgttgtt ggggttttc cataggctc gccccctga 2520 ataccagagg tttccccctg gaagtcccct ggtgggctc cctgttcga gactataaag 2700 ataccagggg tttccccctg gaagtccct cgtggggtct cctgttcga cccggttg ttccccctg cgtgaggta cccagtaac ctgtccgcct ttctcccttc gggaagcggg gcctttctc atagctcacg 2820 ctgtaggtat ctcagttcgg gggcttttc cggtaagcagg gccgttagg gcgctttatc gggaagcagg cccgttcag ccggtcggt gggagggggggggg				-		-	
gtgggaagaa attaaagcaa gataggctat taagtgcaga gggagagaaa attgcctccaa 2400 catgtgagga agtaatgaga gaaatcatag aatttcttcc gcttcctcgc tcactgactc 2460 gctgcgctcg gtcgtccgc tgcggcagc ggtatcagct cactcaaagg ggtaatcagc ggtaatcacc gaatcagggg ataacgcagg aaagaacatg tgagcaaaaa ggcagcaaaaa 2580 ggccaggaac cgtaaaaagg ccgcgttgct ggcgtttttc cataggctcc gaccccctga 2640 cgagcatcac aaaaatcgac gctcaaggtaa gaggtggcga aaacccaaggcg tttccccctg gaagctccct cgtgcgcttc cctgttccga ccctgccgct 2760 taccggatac ctgtccgct ttctcccttc gggaagcgtg gcgctttcc atagctcacg 2820 ctgtaggtat ctcagttcgg tgtaggtcgt tcgctccaag ctgggctgtg cccagtcag cccggttcag cccggttcag cccggttcag cccggttcag cccggttcag cccggttcag cccggttcag cgcgcttatc cggtagactac ttgtagggcgt gccttatc cggtaactat cgcttgagt cacaccggt ggccttatc cggtagactac aggattagga gagcagaga 3000 tgtagggcggt gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaagaac 2840 agaacacagac ttatcgccac tggcaggagc cactggtaac aggattagac gagaggaggt gctacagagt tcttgaagtg ggtgcctaac tacggctaca ctagaagaac 3060 agtatttggt atctgcgctc tgctgaaggc aggttacttc ggaaaaaagag ttttgttgaa aaaaaaggag tctttgaac ggtggtttt tttgtttgaa agacagagat 1880 tacggcaga aaaaaaaggat ctcaagaaga ttttggtcatg agactaaaa 3060 cacctagatc cttttaaatt aaaaatgaag ttttggtcatg agattacaa aaaaggagtct gaaaaactcac gttaagggat ttttggtcatg agattacaa aaaaggagtct gacagttaac aatggttaac aatggttaca aatggttaac aatggttaac aatggaggagggggggggg							
catgtgagga agtaatgaga gaaatcatag aatttettee getteetege teactgacte getgegeteg tyegggegage ggtateaget cacteaaagg gecagcaaaa 2580 ggecaggaac cgtaaaaagg cegegttget ggegtttte cataggete gececettga 2520 gtateacgag aaaaaacgag gaggtggega aaaccagag gecagcaaaa 2580 ggecaggaac cgtaaaaagg cegegttget ggegtttte cataggete gececettga 2640 caggcatcac aaaaatcgac gaggteete ggegtttte cettgteega 2760 ataccaggegt ttteecectg gaageteete teeteete gggaageggg gegetteete atacggatac etgteegeet teeteete gggaageggg gegetteete ataggteage etgtaggteg tegetteete ataggteage ceceggteag eegeettate eggtaactat eggegetty tyeacgaace 2880 cecegtteag eeggeettate eggtaactat eggettiggg gaggagggg gaaaaacagac tatategeea tettgaagtg gegeettate eggtaactat eggettiggg gaggagggggggggggggggggggggggggggg							
gctgctcg gtcgttcggc tgcgcgagc ggtatcagct cactcaaagg cggtaatacg gttatccaca gaatcagggg ataacgcagg aaaagacatg tgagcaaaaag gccagcaaaa 2580 ggccaggaac cgtaaaagg ccggttgtt ggcgtttttc cataggctc gccccctga 2640 caaccaggcg tttccccctg gaagctcca cgtggcgtc cctgttccga ccctgcgct tccccctg gaagctcca cgtggcgtct cctgttccga ccctgccgct tcccccttc gggaagcgg acccggataca ctcagttcgg tgaggtcgt tcgcccaagtc ggcgctttctc atagctcacg 2820 cccgttcag cccgaccgct tctcccttc gggaagcgtg gcgctttctc atagctcacg 2840 cccgttcag cccgaccgct gcgccttatc cggtaactat cggtctgggt tcgcacgaacc 2840 aagacacgac ttatcgccac tggcagcacg cactggtaac aggattagca gagcgaggta 3000 tgtaggcggt gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaagac 2940 agaacacgac tattcgcac tggcagcacg cactggtaac aggattagca gagcgaggta 3000 tgtaggcggg aaaaacaca ccgctggtag cggtgttt tttgttgca gagcgaggta 3180 tacgcgcaga aaaaaaggat ctcaagaaga tcttagatc ggtagcttc tttttaaatt aaaaatggaag ttttgatca accctaggtc gaaaacccac ggttaatca cacttggtca aaaaaccac ccgctggaa gggggggggg							
gttatccaca gaatcagggg ataacgagg aaagaacatg tgagcaaaaa gccagcaaaa ggccaggaac cgtaaaaagg cgcgttgct ggggtttttc cataggctcc gccccctga 2640 cgagcatcaa aaaaatcgac gatcaagtca gaggtggga aacccgacag gactataaag 2700 ataccaggcat ttcccccttg gaagctccct cgtgcgctct cctgttccga cctgtccgct tctccccttc gggaagcggg gcgctttctc atagctcacg 2820 cccgttcag cccgaccgc tgtaggtcgt tcgctccaag cgggaagcgtg gcgctttctc atagctcacg 2820 cccgttcag cccgaccgc gcgccttatc cggtaactat cggtcttgat tgaccgaacc 2880 cccgatcagg gctacaagagt tcttgaagtg gtgcctaac aggattagca gagcgaggta atctgagggg gctacaagagt tcttgaagtg gtggcctaac aggattagca gagcgaggta 3000 tgtagggcgg gctacaagagt tcttgaagtg gtggcctaac atcggctaca ctagaagaac 3060 agtatttggt atctgcgct tgctgaagcc agttaccttc ggaaaaagag ttggtagctc 3120 tagtgggaac gaaaactcac ccgctggtag cggtgtttt tttgttcaa agcagcagat 3180 taagtggaac aaaaaaggat ctttaaagag tctttgatc tttttcacag ggctctgacg 23240 cactaggac aaaaactcac gttaagggat ttttggtcatg agattatcaa aaaggatctt 3300 cacctagatc cttttaaatt aaaaatgaag ttttggtcatg agattatcaa aaaggatctt 3300 cacctagatc cttttaaatt aaaaatgaag ttttggtcatg agattatcaa aaaggatctt 3300 gaagccacgg ttgatgag ccttattaat cagtgaggca cctatctcag gggggggggg	catgtgagga	agtaatgaga	gaaatcatag	aatttcttcc	gcttcctcgc	tcactgactc	2460
gttatccaca gaatcagggg ataacgagg aaagaacatg tgagcaaaaag gccagcaaaa 2580 ggccaggaac cgtaaaaagg ccgcgttgct ggcgtttttc cataggctc gccccctga 2640 ataccaggcg tttececetg gaagctccet cgtgcgctct cctgttccga ccctgccgct tcctccqcct ttctccctc gggaagcgtg gcgctttctc atagctcacg ctgtaggtat ctcagttcgg tgtaggtgt tcgctccaag ctgggctgt tgcaccacgg 2820 cccgttcag cccgaccgct tgtaggtgt tcgctcaaag ctgggctgt tgcacgaacc 2880 cccgttcag cccgaccgct tgtaggtagt cggtaactat cggtataca aggattagca gacgaggta 3000 tgtaggggg gctacaagagt tcttgaaggt gtgggctaac aggattagca gacgaggta accacggt accaggagat accaggaga accaaaacca tgcggaacaaaagga tcctaaagaag tcctaaagaag tttggtagccaaaaggagaaaaaaggaga cccaagagaa tctttggta accacagaga tcctaaagaaga tcctaaaaaggaa tcttggtagaac gaaaaactcac cacctagatc aaaaatgaag tttggtaagca aaaaatgaag tttggtagaac aaaaatgaag tttggtagaac aaaaatgaag tttggtagaac aaaaatgaag tttggtagaac cactaggtaac aatgcttaat cagtgaggagggggggggg	gctgcgctcg	gtcgttcggc	tgcggcgagc	ggtatcagct	cactcaaagg	cggtaatacg	2520
ggccaggaac cgtaaaaagg ccgcgttgct ggcgtttttc cataggctcc gccccctga 2760 aaaacaggcg tttccccctg gaagetccct cgtgcgctct cctgtctcga ccctgccgct tccccggatac ctgtccgcct ttctcccttc gggaagegtg gcgctttctc atagctcacg 2820 ctgtaggtat ctcagttcgg tggaggtggt tcgctcaag ctgggctgt tgcacgaacc 2880 ccccgttcag cccgaccgct gcgccttatc cggtaactat cgtcttgagt ccaacccggt aagacacgac ttatcgccac tgcaggacg cactggtaac aggattagca gagggaggt ggcacaagagt tcttgaagtg gcgcacaacaagac tctggaagga gcacaagagt tctgaaggag gcacaagagt tgctgaagag gggggggggg							2580
cgagcatcac aaaaatcgac gctcaagtca gaggtggcga aacccgacag gactataaag 2700 ataccaggcg tttccccctg gaagctccct cgtgcgctct cctgttccga ccctgccgct 2760 taccggatac ctgtccgcct ttctcccttc gggaagcgtg gcgctttctc atagctcacg 2820 ctgtaggtat ctcagttcgg tgtaggtcgt tcgctccaag ctgggcgttg tgcacgaacc 2880 ccccgttcag cccgaccgct gcgccttatc cggtaactat cgtcttgagt ccaacccggt agagacacgac ttatcgcac tggcagcag cactggtaac aggattagca gagcgaggta 3000 tgtaggcggt gctacaagat tcttgaagtg gtggcctaac acggctaca ctagaaagaac agtatttggt atctgcgcc tgctgaagcc agttaccttc ggaaaaaagaac ttggtagctc ttgatccggc aaaaaaacac ccgctggtag cggtggtttt tttgttgaa agcagcagat tacggcgaa aaaaaaggat cctaagaaga tcctttgatc ttttctacag ggtctgacca agcagcagat tttggtagctc accaggagaac agttacctac ggaaaaaagag ttggtagctc 3120 tacggcgaa aaaaaaaggat ctcaaagaaga tcctttgatc ttttctacagg ggtctgacgc 3240 caactagatc atttaaatt aaaaaatgaag ttttaaaatca atctaaaggat ttttggtcatg aggttatcaa aaaaggatctt aaacttggtct gacagttacc aatgcttaat cagtgaggca cctaaccaag cgatctgct aacttggtct gacagttacc aatgcttaat cagtgaggca cctaaccaag cgatctgct aacttggtcg gagagcacgg ttggaccag ttggatgaga cctttgttga gggggggggg	ggccaggaac	cotaaaaaaa	ccacattact	ggcgtttttc	cataggetee	accccctaa	
ataccaggcg tttcccctg gaagctcct cgtgcgctct cctgttccga ccctgccgct taccggatac ctgtccgct ttctcccttc gggaagcgtg gcgctttctc atagctcacg 2820 ctgtaggtat ctcagttcgg tgtaggtcgt tcgctccaag ctgggcgttgt tgcacgaacc 2880 ccccgttcag cccgaccgct gcgccttatc cggtaactat cgtcttgagt ccaacccggt 2940 aagacacgac ttatcgccac tggcagcagc cactggtaac aggattagca gagcgaggta 3000 tgtaggcggt gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaagaac 3060 agtatttggt atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagccc 3120 ttgatccggc aaacaaacca ccgctggtag cggtggtttt tttgtttgca agcagcagat 3180 tacgcgcaga aaaaaaggat ctcaagagat tcttggtcatg agattaccaa ggattagcacg ggtcagagga cactggtaac atctttaaat aaaaatgaag ttttggtcatg ggttctaacg ggtctgacgc 3240 aacttggtct gacagttacc aatgcttaat cagtgaggca cctatctcag cgatctgtct 3420 aacttggtct gacagttacc aatgcttaat cagtgaggca cctatctcag cgatctgtct 3420 atttcgtca tccatagttg cctgactcgg gggggggggg							
taccggatac ctgtccgct ttctcccttc gggaagcgtg gcgctttctc atagctcacg ctgtaggtat ctcagttcgg tgtaggtcgt tcgctccaag ctgggctgtg tgcacgaacc 2880 ccccgttcag cccgaccgct gcgccttatc cggtaactat cgtcttgagt ccaacccggt 2940 aagacacgac ttatcgccac tggcagcagc cactggtaac aggattagca gagcgaggta 3000 tgtaggcggt gctacagagt tcttgaagtg gtggcctaac acggctacac ctagaagaac 3060 agtatttggt atctgcgctc tgctgaagcc agttaccttc ggaaaaaagag ttggtagctc ttgatgcgga aaaaaaaggat ctcaagaaga tcctttgatt tttgtttgca agcagcagat 3180 tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc 3240 taggtggaac gaaaactcac gttaagggat tttggtcatg agattatcaa aaaggatctt 3300 cacctagatc cttttaaatt aaaaatgaag ttttaaatca atctaaagta tatatgagta 3360 aacttggtct gacagttacc aatgcttaat cagtgaggca cctatctcag cgatctgtct attcctatgttg cctgactcgg gggggggggg							
ctgtaggtat ctcagttcgg tgtaggtcgt tcgctcaag ctgggctgtg tgcacgaacc ccccgttcag cccgaccgct gcgccttatc cggtaactat cgtcttgagt ccaacccggt 2940 aagacacgac ttatcgcac tggcagcag cactggtaac aggattagca gagcgaggta 3000 tgtagggcggt atctgcaggt ctcttgaagtc gtggcctaac tacggctaca ctagaagaac 3060 agtatttggt atctgcgct tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccggc aaacaaacca ccgctggtag cggtggtttt tttgttgca agcagcagat 3120 tacggcgaga aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc 3240 taagtggaac gaaaactcac gttaagggat tttggtcatg agattatcaa aaaggatctt aacttggtct gacagttacc aattggtca gacagttacc aattggtcag gggggggggg							
ccccgttcag cccgaccgct gcgccttatc cggtaactat cgtcttgagt ccaaccggt agacacgac ttatcgccac tggcagcagc cactggtaac aggattagca gagcgaggta 3000 tgtaggcggt gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaagaac 3060 agtatttggt atctgcgctc tgctgaagcc agttaccttc ggaaaaaagag ttggtagctc 3120 tacggcaga aaaaaaggat cccaagaaga tcctttgatc tttgtttgca agccagcagat 3180 tacggcaga aaaaaaggat ctcaagaaga tctttgatc ggtctgacgc ggtctgacgc gttagtggaac gaaaactcac gttaaagggat ttttgtcatg ggtctgacgc 3240 tacgtggaac gaaaactcac gttaaagggat ttttgtcatg ggtctgacgc 3360 aacttggtct gacagttacc aatgcttaat cagtggaggca cctatctcag ggtctgacgt 3360 aacttggtct gacagttacc aatgcttaat cagtggaggca cctatctcag cgatctgtc 3420 agaaggtgtt gctgactcat accaggcctg gggggggggg							
aagacacgac ttatcgcac tggcagcagc cactggtaac aggattagca gagcgaggta 3000 tgtaggcggt gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaagaac 3060 agtatttggt atctgcgcc tgctgaagcc agttaccttc ggaaaaagagg ttggtagctc ttgatccggc aaacaaacca ccgctggtag cggtggtttt tttgttgca agcagcagat 3180 tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc 3240 cactaggtc gaaaactcac gttaagggat ttttggtcat agattatcaa aaaggatctt aaaattggtct gacagttacc aaatggctaat cagtgaggca cctatctcag cgatctgtct atttcgttca tccatagttc aatgcttaat cagtgaggca cctatctcag cgatctgtct atttcgttca tccatagttc accaggcctg gggggggggg	ctgtaggtat	ctcagttcgg	tgtaggtcgt	tcgctccaag	ctgggctgtg	tgcacgaacc	
tgtaggeggt gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaagaac 3060 agtatttggt atctgcgcte tgctgaagce agttacctte ggaaaaagag ttggtagcte 3120 tacgcgcaga aaaaaaggat ctcaagaaga tcctttgate ttttgttgaa agcagcagat 3180 tacgcgcaga aaaaaaggat ctcaagaaga tcctttgate ttttetacgg ggtctgaege 3240 tacgtggaac cttttaaatt aaaaatgaag ttttggtcatg agattatcaa aaaggatett 3300 aaacttggtet gacagttace aatgcttaat cagtggaggag cctatetcag cgatetgte 3300 aatttggtet gacagttace aatgcttaat cagtggaggagggggggggggggggggg	ccccgttcag	cccgaccgct	gcgccttatc	cggtaactat	cgtcttgagt	ccaacccggt	2940
tgtaggcggt atctgaagt tcttgaagtg gtggcctaac tacggctaca ctagaagaac 3060 agtatttggt atctgcgcc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc 3120 ttgatccggc aaacaaacca ccgctggtag cggtggtttt tttgttgca agcagcagat 3180 tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc 3240 gaaacttcac gttaagggat tttggtcatg agattatcaa aaaggatctt 3300 aaacttggtct gacagttacc aatgcttaat cagtgagggagggggggggg	aagacacgac	ttatcgccac	tggcagcagc	cactggtaac	aggattagca	gagcgaggta	3000
agtatttggt atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc 3120 ttgatccggc aaacaaacca ccgctggtag cggtggtttt tttgtttgca agcagcagat 3180 tacgcgcaga aaaaactcac gttaagggat tcctttgatc ttttctacgg ggtctgacgc 3240 cacctagatc cttttaaatt aaaaatgaag ttttggtcatg agattatcaa aaaggatctt 3300 cacctagatc cttttaaatt aaaaatgaag ttttaaatca atctaaagta tatatgagta 3360 aacttggtct gacagttacc aatgcttaat cagtgaggggggggg cgctgaggtc tgcctggtag 3480 agaaggtgtt gctgactcat accaggcctg aatcgccca tcatccagc agaaagtgag 3540 ggagccacgg ttgatgagag ctttgttgta ggtggaccag ttggtgattt tgaacttttg 3600 ctttgccacg gaacggtctg cgttgtcgg aagatgcgtg atctgatct tcaactcagc 3660 aaaagttcga tttattcaac aattaaccaa ttctgattag aaaaactcat cgagcatcaa atgaaactgc 3780 aattattac tatcaggat accagagca gtccatagg atggaaaact caccgaggca gtccatag atggcaagaa ccttggtatca 3900 aggagaaaact caccatgagt gacgactga tccggtgaga atggcaaaac cctggattaca aggccagcca ttatgccct tcaacaca 4080							3060
ttgatcegge aaacaaacca cegetggtag eggtggtttt tttgtttgea ageageagat 3240 taegegeaga aaaaaaggat eteetttgate ttttetaegg ggtetgaege 3240 teagtggaac gaaaacteae gttaagggat tttggteatg agattateaa aaaggatett 3300 caectagate etettaaatt aaaaatgaag ttttaaatea atetaaagta tatatgagta 3360 aacttggtet gaeagttaee aatgettaat eagtgaggea eetateteag egatetgtet 3420 atttegttea teeatagttg eetgaetegg gggggggggg							3120
tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc gaaaactcac gttaagggat tttggtcatg agattatcaa aaaggatctt 3300 cacctagatc cttttaaatt aaaaatgaag ttttaaatca atctaaagta tatatgagta 3360 aacttggtct gacagttacc aatgcttaat cagtgaggca cctatctcag cgatctgtct atccatagttg cctgactcgg gggggggggg							
tcagtggaac gaaaactcac gttaagggat tttggtcatg agattatcaa aaaggatctt cacctagatc cttttaaatt aaaaatgaag ttttaaatca atctaaagta tatatgagta 3360 aacttggtct gacagttacc aatgcttaat cagtgaggca cctatctcag cgatctgtct 3420 atttcgttca tccatagttg cctgactcgg gggggggggg							
cacctagatc cttttaaatt aaaaatgaag ttttaaatca atctaaagta tatatgagta 3360 aacttggtct gacagttacc aatgcttaat cagtgaggca cctatctcag cgatctgtct 3420 atttcgttca tccatagttg cctgactcgg gggggggggg	tacgcgcaga	aaaaaaggac	cucaayaaya	tectityate	cuttudadgg	ggtctgacgt	
aacttggtct gacagttacc aatgcttaat cagtgaggca cctatctcag cgatctgtct 3420 atttegttca tecatagttg cctgactegg gggggggggg cgctgaggtc tgcctcgtga 3480 agaaggtgtt gctgactcat accaggcctg aategeccca tcatccagcc agaaaggtgag gggggcacagg ttggtgactgg ctttgttgta ggtggaccagg ttggtgattt tgaacttttg 3600 ctttgccacg gaacggtctg cgttgtcggg aagatgcgtg atctgatcct tcaactcagc 3660 aaaagttcga tttattcaac aattaaccaa ttctgatag aaaaactcaat cgagcatcaa atgaaactgc 3720 tgttacaacc aattaaccaa ttctgatag aaaaactcaat cgagcatcaa atgaaactgc 3780 agtgagaaaact caccgaggca gttccatagg atggcaagat cctggtatcg gtctgcgatt 3900 ccgactcgtc caacactcaat acaacctatt aatttcccct cgtcaaaaat aaggttatca 3960 agtgagaaaat caccatgagt gacgactgaa tccggtgaga atggcaaaag cttatgcatt 4020 tctttccaga cttgttcaac aggccagcca ttacgctcgt catcaaaatc actcgcatca 4080							
atttegttea tecatagttg cetgaetegg ggggggggg egetgaggte tgeetegtga 3480 agaaggtgtt getgaeteat aceaggeetg aategeecea teatecagee agaaagtgag 3540 ggageeaegg ttgatgagag etttgttgta ggtggaeeagg ttggtgattt tgaaettttg 3600 etttgeeaeg gaaeggtetg egttgteggg aagatgegtg atetgateet teaaeteage 3660 aaaagttega tttatteaae aaageegeegg teeeggteaaggeggtaat getetgeeag 3720 tgttacaaee aattaaeeaa teeaataeea tattttgaa aaageeggtt etgaaagae 3780 aattattea tateaggatt ateaataeea tatttttgaa aaageeggtt etgaatgaa 3840 ggagaaaaet eacegaggea gtteeatagg atggeaagae eetggtateg gtetgegatt 3900 aggaetegte eaceateaat acaaeetatt aattteeet teaaeteea 3780 aggagaaaat eacegaggea gtteeatagg atggeaagae eetggtateg gtetgegatt 3900 eetggaetege eaceategat acaaeetatt aattteeet egteaaaaat aaggttatea 3960 agtgagaaat eaceatgagt gaegaetgaa teeggtgaga atggeaaaag ettatgeatt 4020 tettteeaga ettgtteaae aggeeageea ttaegetegt eateaaaate aetegeatea 4080	cacctagatc	cttttaaatt	aaaaatgaag	ttttaaatca	atctaaagta	tatatgagta	
agaaggtgtt gctgactcat accaggcctg aatcgcccca tcatccagcc agaaagtgag ggagccacgg ttgatgagag ctttgttgta ggtggaccag ttggtgattt tgaacttttg 3600 ctttgccacg gaacggtctg cgttgtcggg aagatgcgtg atctgatcct tcaactcagc 3660 aaaagttcga tttattcaac aaagccgccg tcccgtcaag tcagcgtaat gctctgccag 3720 tgttacaacc aattaaccaa ttctgattag aaaaactcat cgagcatcaa atgaaactgc 3780 aatttattca tatcaggat atcaatacca tatttttgaa aaagccgttt ctgtaatgaa 3840 ggagaaaact caccgaggca gttccatagg atggcaagat cctggtatcg gtctgcgatt 3900 agtgagaaat caccatgagt gacgactgaa tccggtgaga atggcaaaag cttatgcatt 4020 tctttccaga cttgttcaac aggccagcca ttacgctcgt catcaaaatc actcgcatca 4080	aacttggtct	gacagttacc	aatgcttaat	cagtgaggca	cctatctcag	cgatctgtct	3420
ggagccacgg ttgatgagag ctttgttgta ggtggaccag ttggtgattt tgaacttttg 3600 ctttgccacg gaacggtctg cgttgtcggg aagatgcgtg atctgatcct tcaactcagc 3660 aaaagttcga tttattcaac aaagccgccg tcccgtcaag tcagcgtaat gctctgccag 3720 tgttacaacc aattaaccaa ttctgattag aaaaactcat cgagcatcaa atgaaactgc 3780 aatttattca tatcaggatt atcaatacca tattttgaa aaagccgttt ctgtaatgaa ggagaaaact caccgaggca gttccatagg atggcaagat cctggtatcg gtctgcgatt 3900 agtgagaaat caccatgagt gacgactgaa tccggtgaga atggcaaaag cttatgcatt 4020 tctttccaga cttgttcaac aggccagcca ttacgctcgt catcaaaatc actcgcatca 4080	atttcgttca	tccatagttg	cctgactcgg	ggggggggg	cgctgaggtc	tgcctcgtga	3480
ggagccacgg ttgatgagag ctttgttgta ggtggaccag ttggtgattt tgaacttttg 3600 ctttgccacg gaacggtctg cgttgtcggg aagatgcgtg atctgatcct tcaactcagc 3660 aaaagttcga tttattcaac aaagccgccg tcccgtcaag tcagcgtaat gctctgccag 3720 tgttacaacc aattaaccaa ttctgattag aaaaactcat cgagcatcaa atgaaactgc 3780 aatttattca tatcaggatt atcaatacca tattttgaa aaagccgttt ctgtaatgaa ggagaaaact caccgaggca gttccatagg atggcaagat cctggtatcg gtctgcgatt 3900 agtgagaaat caccatgagt gacgactgaa tccggtgaga atggcaaaag cttatgcatt 4020 tctttccaga cttgttcaac aggccagcca ttacgctcgt catcaaaatc actcgcatca 4080	agaaggtgtt	gctgactcat	accaggeetg	aatcgcccca	tcatccagcc	agaaagtgag	3540
ctttgccacg gaacggtctg cgttgtcggg aagatgcgtg atctgatect tcaactcage 3660 aaaagttcga tttattcaac aaagccgccg tcccgtcaag tcagcgtaat gctctgccag 3720 tgttacaacc aattaaccaa ttctgattag aaaaactcat cgagcatcaa atgaaactgc 3780 aatttattca tatcaggatt atcaatacca tatttttgaa aaagccgttt ctgtaatgaa 3840 ggagaaaact caccgaggca gttccatagg atggcaagat cctggtatcg gtctgcgatt 3900 ccgactcgtc caacatcaat acaacctatt aatttcccct cgtcaaaaat aaggttatca 3960 agtgagaaat caccatgagt gacgactgaa tccggtgaga atggcaaaag cttatgcatt 4020 tctttccaga cttgttcaac aggccagcca ttacgctcgt catcaaaatc actcgcatca 4080							3600
aaaagttega tttatteaac aaagcegeeg teeegteaag teagegtaat getetgeeag 3720 tgttacaace aattaaceaa ttetgattag aaaaacteat egageateaa atgaaactge 3780 aatttattea tateaggatt ateaatacea tatttttgaa aaagcegttt etgtaatgaa 3840 ggagaaaact caccgaggea gttecatagg atggeaagat eetggtateg gtetgegatt 3900 eegactegte eaacateaat acaacetatt aattteeet egteaaaaat aaggttatea 3960 agtgagaaat eaceatgagt gacgactgaa teeggtgaga atggeaaaag ettatgeatt 4020 tettteeaga ettgtteaac aggeeageea ttaegetegt eateaaaate aetegeatea 4080							
tgttacaacc aattaaccaa ttctgattag aaaaactcat cgagcatcaa atgaaactgc 3780 aatttattca tatcaggatt atcaatacca tatttttgaa aaagccgttt ctgtaatgaa 3840 ggagaaaact caccgaggca gttccatagg atggcaagat cctggtatcg gtctgcgatt 3900 ccgactcgtc caacatcaat acaacctatt aatttcccct cgtcaaaaat aaggttatca 3960 agtgagaaat caccatgagt gacgactgaa tccggtgaga atggcaaaag cttatgcatt 4020 tctttccaga cttgttcaac aggccagcca ttacgctcgt catcaaaatc actcgcatca 4080							
aatttattca tatcaggatt atcaatacca tatttttgaa aaagccgttt ctgtaatgaa 3840 ggagaaaact caccgaggca gttccatagg atggcaagat cctggtatcg gtctgcgatt 3900 ccgactcgtc caacatcaat acaacctatt aatttcccct cgtcaaaaat aaggttatca 3960 agtgagaaat caccatgagt gacgactgaa tccggtgaga atggcaaaag cttatgcatt 4020 tctttccaga cttgttcaac aggccagcca ttacgctcgt catcaaaatc actcgcatca 4080							
ggagaaaact caccgaggca gttccatagg atggcaagat cctggtatcg gtctgcgatt 3900 ccgactcgtc caacatcaat acaacctatt aatttcccct cgtcaaaaat aaggttatca 3960 agtgagaaat caccatgagt gacgactgaa tccggtgaga atggcaaaag cttatgcatt 4020 tctttccaga cttgttcaac aggccagcca ttacgctcgt catcaaaatc actcgcatca 4080							
ccgactcgtc caacatcaat acaacctatt aatttcccct cgtcaaaaat aaggttatca 3960 agtgagaaat caccatgagt gacgactgaa tccggtgaga atggcaaaag cttatgcatt 4020 tctttccaga cttgttcaac aggccagcca ttacgctcgt catcaaaatc actcgcatca 4080							
agtgagaaat caccatgagt gacgactgaa tccggtgaga atggcaaaag cttatgcatt 4020 tctttccaga cttgttcaac aggccagcca ttacgctcgt catcaaaatc actcgcatca 4080	ggagaaaact	caccgaggca	gttccatagg	atggcaagat	cctggtatcg	gtctgcgatt	3900
agtgagaaat caccatgagt gacgactgaa tccggtgaga atggcaaaag cttatgcatt 4020 tctttccaga cttgttcaac aggccagcca ttacgctcgt catcaaaatc actcgcatca 4080	ccgactcqtc	caacatcaat	acaacctatt	aatttcccct	cgtcaaaaat	aaggttatca	3960
tetttecaga ettgtteaac aggeeageea ttacgetegt cateaaaate actegeatea 4080							
accasactyr tarreattry tyarrycytt tydycydyae ysaatacyty attyctytta 4140							
	uccaaaccyt	carreating	cyarrycyce	Lyaycyayac	gaaacacycy	accyclycia	4140

```
aaaggacaat tacaaacagg aatcgaatgc aaccggcgca ggaacactgc cagcgcatca
                                                                      4200
acaatatttt cacctgaatc aggatattct tctaatacct ggaatgctgt tttcccgggg
                                                                      4260
atcgcagtgg tgagtaacca tgcatcatca ggagtacgga taaaatgctt gatggtcgga
                                                                      4320
agaggcataa attccgtcag ccagtttagt ctgaccatct catctgtaac atcattggca
                                                                      4380
acgctacctt tgccatgttt cagaaacaac tctggcgcat cgggcttccc atacaatcga
                                                                      4440
tagattgtcg cacctgattg cccgacatta tcgcgagccc atttataccc atataaatca
                                                                      4500
gcatccatgt tggaatttaa tcgcggcctc gagcaagacg tttcccgttg aatatggctc
                                                                      4560
ataacacccc ttgtattact gtttatgtaa gcagacagtt ttattgttca tgatgatata
                                                                      4620
tttttatctt gtgcaatgta acatcagaga ttttgagaca caacgtggct ttccccccc
                                                                      4680
ccccattatt gaagcattta tcagggttat tgtctcatga gcggatacat atttgaatgt
                                                                      4740
atttagaaaa ataaacaaat aggggttccg cgcacatttc cccgaaaagt gccacctgac
                                                                      4800
gtctaagaaa ccattattat catgacatta acctataaaa ataggcgtat cacgaggccc
                                                                      4860
                                                                      4867
tttcgtc
<210> 17
<211> 75
<212> DNA
<213> Artificial Sequence
<220>
<223> oligonucleotide
<400> 17
gatcaccatg gatgcaatga agagagggct ctgtgtgctg ctgctgtgtg gagcagtctt
                                                                        60
                                                                        75
cgtttcgccc agcga
<210> 18
<211> 78
<212> DNA
<213> Artificial Sequence
<220>
<223> oligonucleotide
gatctcgctg ggcgaaacga agactgctcc acacagcagc agcacacagc agagccctct
                                                                       . 60
                                                                        78 -
cttcattgca tccatggt
<210> 19
<211> 33
<212> DNA
<213> Artificial Sequence
<220>
<223> oligonucleotide
                                                                        33
ggtacaaata ttggctattg gccattgcat acg
<210> 20
<211> 36
<212> DNA
<213> Artificial Sequence
<220>
<223> oligonucleotide
<400> 20
                                                                        36
ccacatctcg aggaaccggg tcaattcttc agcacc
<210> 21
<211> 38
```

<212> DNA <213> Artificial Sequence	•
<220> <223> oligonucleotide	
<400> 21 ggtacagata tcggaaagcc acgttgtgtc tcaaa	atc 38
<210> 22 <211> 36 <212> DNA <213> Artificial Sequence	
<220> <223> oligonucleotide	
<400> 22 cacatggatc cgtaatgctc tgccagtgtt acaac	ec . 36
<210> 23 <211> 39 <212> DNA <213> Artificial Sequence	
<220> <223> oligonucleotide	
<400> 23 ggtacatgat cacgtagaaa agatcaaagg atctt	acttg 39
<210> 24 <211> 35 <212> DNA <213> Artificial Sequence	
<220> <223> oligonucleotide	
<400> 24 ccacatgtcg acccgtaaaa aggccgcgtt gctgg	
<210> 25 <211> 4864 <212> DNA <213> E. coli (V1R plasmid)	
<pre><400> 25 tcgcgcgttt cggtgatgac ggtgaaaacc tctgacagcttgtct gtaagcggat gccgggagga gacaattggggggggt tcggggctgg cttaactatg cggcacattgggggggggg</pre>	gcccg tcagggcgcg tcagcgggtg 120 tcaga gcagattgta ctgagagtgc 180 aggag aaaataccgc atcagattgg 240 atatg tacatttata ttggctcatg 300 actagt tattaatagt aatcaattac 360 gcgtt acataacta cggtaaatgg 420 tagacg tcaataatga cgtatgttcc 480 taatgg gtggagtatt tacggtaaac 540 acaagt acgccccta ttgacgtcaa 600 accatg accttatggg actttcctac 660 accatg gtgatgcggt tttggcagta 720
cgtcaatggg agtttgtttt ggcaccaaaa tcaac	

				hagaaatat	atataadcad	900
ctccgcccca	ttgacgcaaa	tgggcggtag	gcgtgtacgg	agaggaggagtatt	ttgacctcca	960
agctcgttta	gtgaaccgtc	agatcgcctg	gagacgccat	geatgatta	gaccccca	1020
tagaagacac	cgggaccgat	ccagcctccg	cggccgggaa	tataggggg	ccccttaac	1080
tccccgtgcc	aagagtgacg	taagtaccgc	ctatagagtc	anagagagat	tcctcatatt	1140
ttcttatgca	tgctatactg	tttttggctt	ggggtctata	gaccattatt	gaccactccc	1200
ataggtgatg	gtatagctta	gcctataggt	gtgggttatt	gaccaccacc	acaactctct	1260
ctattggtga	cgatactttc	cattactaat	ccataacatg	geceeectet	dtatttttac	1320
ttattggcta	tatgccaata	cactgtcctt	cagagactga	accegacce	tccccaatac	1380
aggatggggt	ctcatttatt	atttacaaat	tcacatatac	atctcactg	catattccaa	1440
ccgcagtttt	tattaaacat	aacgtgggat	ctccacgcga	gaccctgct	cccatacctc	1500
acatgggctc	ttctccggta	gcggcggagc	ttctacatcc	gageeeegee	gacttaggca	1560
cagcgactca	tggtcgctcg	gcagctcctt	gctcctaaca	ataacaataa	gatatatata	1620
cagcacgatg	cccaccacca	ecagtgtgcc	gcacaaggcc	tttggaagac	ttaaggcagc	1680
tgaaaatgag	ctcggggagc	gggettgeac	cgctgacgca	taagagtcag	aggtaactcc	1740
ggcagaagaa	gatgcaggca	getgagttgt	tgtgttctga tgtagtctga	acagtactca	ttactaccac	1800
cgttgcggtg	ctgttaacgg	cygagygcag	taacagactg	treettteea	tagatettt	1860
gcgcgccacc	agacataata	tatastatas	cttctagttg	ccagccatct	attatttacc	1920
ctgcagtcac	cgtccttaga	congregation	gtgccactcc	cactgtcctt	tcctaataaa	1980
cctccccgt	geetteetty	tatotasata	ggtgtcattc	tattctgggg	aataaaataa	2040
atgaggaaat	tgcatcycat	gettgagea	acaatagcag	gcatgctggg	gatgcggtgg	2100
ggcagcacag	tagggggag	ctcaacaatt	gacccggttc	ctcctaaacc	agaaagaagc	2160
gctctatggg	acttatatat	cegaagaacc	gtccacgccc	ctggttctta	gttccagccc	2220
aggeacatee	ngagtgatag	ctcaccaccc	ctccgccttc	aatcccaccc	gctaaagtac	2280
cacccacagg	atattcatag	ctcatcacc	caccaaacca	aacctagcct	ccaagagtgg	2340
	aaggaagata.	aactattaaa	rocadaddda	yayaaaatyc	Ctccaacacg	2400
gaagaaatta	atrararaa	tcatagaatt	tcttccgctt	cctcgctcac	tgactcgctg	2460
	++~~~~~	acasacaars	Ecaucicaci	Caaayycyyc	aacacggcca	2520
+cananaat	candddataa	cacaaaaaaa	aacatotoao	Caaaayycca	gcaaaaggcc	2300
	22220000000	attactaaca	ELLLLCCALA	ggctccgccc	ccccgacgag	2640
	atomacocto	aagtcagagg	tggcgaaacc	cyacayyacı	acadagacac	2700
	accetaceea	ctccctcata	cacteteeta	Licigation	geegeeeace	2760
toactat	ccacctttct	cccttcaaaa	adcataacac	LLLLLLaatg	cccacgccgc	2820
	atteaatata	antenticae	tccaacciqu	gergegea	cgaaccccc	2880
	accortoror	cttatccccc	aactatcutc	Ligagiccaa	cccggcaaga	2940
	- caccactaac	accadecact	docaacaqqa	LLaycayayc	gaggcacgca	3000
	a a a a a a t t a t t		cctaactacu	gctacactag	aaggacagca	3060
	acactetact	- daadccadt E	accetecudad	aaayayiig	Lugcccccgu	3120
+444441334	2220020000	taataacaat	aatttttu	Litycaayca	gcagaccacg	3180
-~	aannatrtra	agaagaECCE	ttgatcttt	. Clacygygcc	Lydegeeedg	3240
L	actcacctta	agggatttt	r dicaldadal	lattaaaaay	gattttttt	3300
L	+	atgaagtttt	aaatcaatct	. adaytatata	Lyaytaaact	3360
L	. <u>attaccasta</u>	cttaatcagt	daddcaccia	Lucayuyau	cegeeeace	3420 3480
	. teattaccta	actccaaaa	r aaaaaaaacac	Lyayyucuyu	cccgcgaaga	3540
	- dactratacr	aggeergaat	coccccatca	Liciagicaga	aagtgagga	3600
gaaaaaaatta	, atmamamett	Fattataggt	gaaccautty	gigaliliga	accettgee	3660
+~~~~~~~~	. caatctacat	tataaaaa	ratocotualo	Lyaluuluu	acccagcaaa	3720
	· attraaraaa	- מככמככמדכנ	: cottaautta	i guguaatgu	. ccgccagaga	3780
h	· +cc	. taattaaaaa	aactcatcu	Lycattaaaty	aaactgcaat	3840
++-++ + -+	· caccattate	· aataccatat	: EEEEGaaaaa	t decement	Laatgaagga	3900
	, accordant t	· ccataddatc	r dcaada.ccu	, quiallygue	, Lycyactery	3960
actcgtccaa	a catcaataca	acctattaat	traccaragi	. caaaaacaay	gttatcaagt	4020
gagaaatcac	: catgagtgac	gactgaatco	ggtgagaatg	gcaaaagcct	atgcatttct	4080
ttccagactt	gttcaacagg	g ccagccatta	a egetegteat	. caaaaccacc	cgcatcaacc	4140
aaaccgttat	tcattcgtga	ttgcgcctga	gegagaegae	acacycyacc	gctgttaaaa	4200
ggacaattad	: aaacaggaat	: cgaatgcaa	cggcgcagga	a acaccycoay	cgcatcaaca	4260
atattttcac	: ctgaatcago	atattcttct	. datacctygo	acyccyccc	cccggggatc	4320
gcagtggtga	a gtaaccatgo	atcatcagga	a gracygarae	- ctctaacato	ggtcggaaga	4380
ggcataaatt	t ccgtcagcca	a gtttagtct	g accardical	r acttecests	c attggcaacg	4440
ctacctttg	catgtttcag	aaacaactc	r cdadddarogu	, goodcocata	a caatcgatag a taaatcagca	
attgtcgcad	ctgattgcc	gacattates	y cyayectatt	t cccattaaat	taaatcagca atggctcata	4560
tccatgttg	g aatttaatco	g eggeerega	y caagacycu			

```
acaccccttg tattactgtt tatgtaagca gacagtttta ttgttcatga tgatatattt
ttatcttgtg caatgtaaca tcagagattt tgagacacaa cgtggctttc cccccccc
                                                                            4680 ·
cattattgaa gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt
                                                                            4740
tagaaaaata aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc acctgacgtc taagaaacca ttattatcat gacattaacc tataaaaaata ggcgtatcac gaggcccttt
                                                                            4800
                                                                            4860
cgtc
                                                                            4864
<210> 26
<211> 36
<212> DNA
<213> Artificial Sequence
<220>
<223> oligonucleotide
<400> 26
ggtacaagat ctccgcccc atctccccca ttgaga
                                                                             - 36
<210> 27
<211> 33
<212> DNA
<213> Artificial Sequence
<220>
<223> oligonucleotide
<400> 27
ccacatagat ctgcccgggc tttagtcctc atc
                                                                              33
<210> 28
<211> 27
<212> PRT
<213> Homo sapien
<400> 28
Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Cys Gly
                                       10
Ala Val Phe Val Ser Pro Ser Glu Ile Ser Ser
<210> 29
<211> 45
<212> DNA
<213> Artificial Sequence
<220>
<223> oligonucleotide
caggegagat ctaccatggc ccccattagc cctattgaga ctgta
                                                                              45
<210> 30
<211> 48
<212> DNA
<213> Artificial Sequence
<223> oligonucleotide
caggegagat etgeceggge tttaateete ateetgteta ettgecae
                                                                              48
```

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US00/34724

A. CLASSIFICATION OF SUBJECT MATTER IPC(7): A61K 48/00; C12Q 1/70. US CL: 514/44; 435/5; 424/93.1. According to International Patent Classification (IPC) or to both national classification and IPC B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) U.S.: 514/44; 435/5; 424/93.1.						
Documentation	Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched					
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) Medline, embase, scisearch, biosis, caplus and WEST						
C. DOC	UMENTS CONSIDERED TO BE RELEVANT					
Category *	Citation of document, with indication, where ap	propriate, of the relevant passages	Relevant to claim No.			
Y,P	US 6,099,848 A (FRANKEL et al) 08 August 2000	(08.08.2000), page 12 paragraph 6.	1-14, 17			
Ý	WO 97/31115 A2 (MERCK & CO. INC.), 28 Augu	st 1997, page 36.	4			
x	WO 90/10230 A1 (UNIVERSITY OF OTTAWA) 07	7 September 1990, page 11.	17			
Y	US 5,858,646 A (KANG) 12 January 1999 (12.01.19	999), col. 2, lines 10-17	1-14, 17			
Further	r documents are listed in the continuation of Box C.	See patent family annex.				
* S	pecial categories of cited documents:	"T" later document published after the inte date and not in conflict with the appli	ernational filing date or priority cation but cited to understand the			
	t defining the general state of the art which is not considered to be ular relevance	principle or theory underlying the inv "X" document of particular relevance; the				
•	oplication or patent published on or after the international filing date	considered novel or cannot be considered when the document is taken alone	ered to involve an inventive step			
establish specified	t which may throw doubts ou priority claim(s) or which is cited to the publication date of another citation or other special reason (as) It referring to an oral disclosure, use, exhibition or other means	"Y" document of particular relevance; the considered 19 involve an inventive ste combined with one or more other sue being obvious to a person skilled in the	p when the document is h documents, such combination			
· -						
priority (of published prior to the international filing date but later than the date claimed	*&" document member of the same patent				
	actual completion of the international search	Date of mailing of the international sea 09 WAR 2001	arch report			
Name and m	pailing address of the ISA/US		RY J. DEY			
Cor	mmissioner of Patents and Trademarks	BADAI EC	N SDECIALIST			
	x PCT ashington, D.C. 20231	Eleanor Sorbello Telephone No. 703-308-0196 PARALEGAL SPECIALIST TECHNOLOGY CENTER 1600				
	lo. (703)305-3230	Telephone No. 703-308-0196				

INTERNATIONAL SEARCH REPORT

Internat application No.

	PC17US00734724
Box I Observations where certain claims were found unsearchable (Continua	<u></u>
This international report has not been established in respect of certain claims under Artic	le 17(2)(a) for the following reasons:
1. Claim Nos.: because they relate to subject matter not required to be searched by this Au	tthority, namely:
2. Claim Nos.: because they relate to parts of the international application that do not com such an extent that no meaningful international search can be carried out, s	
3. Claim Nos.: 15 & 16 because they are dependent claims and are not drafted in accordance with the 6.4(a).	he second and third sentences of Rule
Box II Observations where unity of invention is lacking (Continuation of Ite	em 2 of first sheet)
This International Searching Authority found multiple inventions in this international app	lication, as follows:
	·
•	
As all required additional search fees were timely paid by the applicant, this searchable claims.	s international search report covers all
 As all searchable claims could be searched without effort justifying an additional fee. 	tional fee, this Authority did not invite
3. As only some of the required additional search fees were timely paid by the report covers only those claims for which fees were paid, specifically claim	
4. No required additional search fees were timely paid by the applicant. Cons is restricted to the invention first mentioned in the claims; it is covered by	
Remark on Protest The additional search fees were accompanied by the appl No protest accompanied the payment of additional search	

Form PCT/ISA/210 (continuation of first sheet(1)) (July 1998)

THIS PAGE BLANK (USPTO)